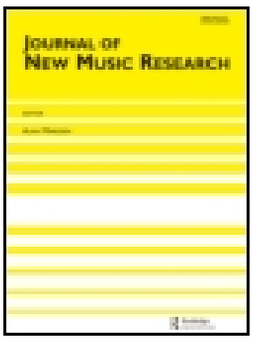

Journal of New Music Research

ISSN: 0929-8215 (Print) 1744-5027 (Online) Journal homepage: https://www.tandfonline.com/loi/nnmr20

Rhythmic variability, language, and style: A replication and extension of nPVI findings with the RISM dataset

Katherine Vukovics & Daniel Shanahan

To cite this article: Katherine Vukovics & Daniel Shanahan (2020): Rhythmic variability, language, and style: A replication and extension of nPVI findings with the RISM dataset, Journal of New Music Research, DOI: 10.1080/09298215.2020.1751209

To link to this article: https://doi.org/10.1080/09298215.2020.1751209

JOURNAL OF NEW MUSIC RESEARCH
https://doi.org/10.1080/09298215.2020.1751209

Routledge
Taylor & Francis Group

Rhythmic variability, language, and style: A replication and extension of nPVI findings with the RISM dataset

Katherine Vukovics[a] and Daniel Shanahan[b]

[a]School of Music, Louisiana State University, Baton Rouge, LA, USA; [b]School of Music, Ohio State University, Columbus, OH, USA

ABSTRACT

The normalised pairwise variability index (nPVI), has been frequently used to measure the rhythmic variance between musical onsets. For example, researchers have proposed connections between the nationalities of composers and the nPVI values of their music and native language. One particular issue, however, lies in the notion of *intended nationality* – composers frequently wrote in national styles other than their own. This study employs the RISM-World dataset to both replicate previous findings relating the rhythmic variance of melodies to the nationalities of composers, and to examine intended nationality in the music of Mozart and Handel, namely their operas and oratorios.

ARTICLE HISTORY
Received 5 January 2019
Accepted 26 March 2020

KEYWORDS
Databases; social aspects; theory; history; music analysis

Introduction

The normalised pairwise variability index (nPVI) was first employed in 2000 by Grabe and Low as a way of measuring the rhythmic variability of linguistic utterances (Grabe & Low, 2002). An utterance with a large contrast between neighbouring durations would be assigned a high nPVI, whereas an utterance with little contrast would have a relatively low nPVI. Grabe and Low argued that the pairwise variability index could be used to identify differences between languages and place them in linguistic groups, such as stress-timed languages (including Dutch, English, and German) and syllable-timed languages (including French and Italian). Patel and Daniele (2003) employed this metric to examine the role of rhythmic variability in musical patterns, mappingnotated rhythms onto the nPVI equation by providing a value for each note based on how its duration related to a single chosen note within the phrase (see Patel & Daniele, 2003, p. B40) For example, a rhythm of nothing but quarter notes contains no variability, and would therefore consist of an nPVI metric of 0, whereas a rhythm that contains longer durations followed by shorter durations would have far more variability (see Patel, 2008, p.164 for a discussion, and Condit-Schultz, 2019, p. 301 for a concise summary). The equation for nPVI is written as follows:

$$nPVI = \frac{100}{m-1} * \sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\left(\frac{d_k + d_{k+1}}{2}\right)} \right| \quad (1)$$

The calculation for the normalised pairwise variability index (nPVI) as proposed by Grabe and Low (2002) and discussed by Daniele and Patel (2004) in Equation (1).

These influential studiesargued for a relationship between a composer's native language and the nPVI ratings of their melodic writing, findings which were later replicated by Huron and Ollen (2003). Recent work has continued this line of examination: Patel et al. (2006) examined the specific cases of British English and French, and argued that pitch transitions could be used to complement the nPVI's predictive power for nationality. Temperley and Temperley (2011) conducted a study in which the 'Scotch snap' – a characteristic rhythm of Scottish music – was found in both linguistic and musical rhythms, and Shanahan and Shanahan (2014) used the nPVI metric to examine the relationship between the songs of Native American populations by linguistic group. A number of studies have also worked to refine the metric. London and Jones (2011), for example, argued that metre should be taken into account (while also suggesting alternative methods for encoding that might take hypermetric variance into account). Toussaint (2012) tested the efficacy of the nPVI metric, comparing it to the standard deviations of inter-onset intervals, and arguing that the metric struggles with non-Western music.

Although the metric has been used primarily as a marker of linguistic origin of composers, the nPVI has also been interpreted as an indicator of explicit stylistic trends in compositional practices. In 2004, Daniele and Patel used the metric to distinguish between the musical

CONTACT Daniel Shanahan shanahan.37@osu.edu

styles at play by delineating between melodies taken from operas, ballets, symphonies, etc. In the years following, the metric was still mainly interpreted within a linguistic context, but recent work has seen a shift towards this idea that the metric might be more aptly used as an indicator of intentional stylistic choices rather than the unconscious influence of native language. Daniele and Patel (2013) argued that the rise of the influence of German nationalism in the eighteenth century might be examined through shifts in rhythmic variability – that is, German composers were intentionally trying to sound 'less Italianate' around the 1780s, at which point the rhythmic variability would begin to diverge. Hansen et al. (2016) tested Daniele and Patel's historically-informed hypothesis, but argued that style is more curvilinear than linear. They employed a polynomial regression in order to replicate their findings. Additionally, they found an increase in the nPVI ratings of French composers, followed by a subsequent decrease. These results provided further support to the idea that the intended compositional style of a composer might also affect the nPVI values. Recently, Daniele (2017b) discussed this idea through the use of rhythmic fingerprints – density plots of the nPVI values of the works of an individual composer. Both he and VanHandel (2017) argued that many of the nPVI ratings can be used to delineate between the two sides of intentional style in the 'War of the Romantics' in the late nineteenth century. The Brahmsian school of composition was considered the more conservative of the two, and the fingerprints of such composers show similarly low nPVI values. The Wagnerians, however, were known to be more unconventional, and their nPVI values seem to reflect that (see Daniele, 2017b, p. 6). The combined evidence from these studies provides strong support for using nPVI as a marker of *stylistic,* rather than *linguistic,* relationships. Some studies have been less enthusiastic of the use of nPVI for the examination of music and language connections. For example, VanHandel (2005) was unable to replicate the findings, despite using vocal pieces in the analysis, and recently, Condit-Schultz (2019) has argued that the nPVI metric should only be used in coordination with other measures.

Despite these extended discussions, many of the studies focused on rhythmic variability repeatedly used the same limited number of datasets. Daniele and Patel (2004, 2013), Huron and Ollen (2003), Patel and Daniele (2003), Patel et al. (2006), Daniele (2017a, 2017b), and Hansen et al. (2016) all used the encodedthemes from the *Barlow and Morgenstern Dictionary of Musical Themes* (Barlow & Morgenstern, 1948). As London (2013) has pointed out, this dataset is not without its flaws: it has a bias toward Germanic composers as well as an overrepresentation from the more prominent ones. Regardless of its validity, however, the repeated use of *any* dataset is problematic, and the broader field of corpus studies is increasingly recognising this reliance on a select few corpora as problematic (see Huron, 2013). If we hope to replicate these findings and provide further evidence towards an analysis of intentionally stylistic gestures of rhythmic variability, it would therefore make sense to do so on a separate dataset. A number of studies did employ the nPVI on other corpora (Shanahan & Shanahan, 2014; Temperley & Temperley, 2011; VanHandel & Song, 2010), but their aims were not to replicate the validity of the measure on a large scale. Our aims here are threefold: first, we hope to replicate the findings of the original work on nPVI with a larger and separate dataset. Secondly, we will examine the possible role of stylistic intent on rhythmic variability, as measured with the nPVI. Lastly, we will discuss how the inclusion of pitch transition probabilities, as discussed in Patel et al. (2006) and Arvaniti (2012) can be a useful metric in examining stylistic differences in incipits.

The dataset

For this study, we used the RISM (Répertoire International des Sources Musicales) dataset (RISM, 2018). The RISM collection of incipits was originally encoded for bibliographic purposes in libraries across the world in the 'Plaine and Easie Code' (a format for the International Association of Music Libraries; see Brook, 1965), and converted into kern notation by Craig Sapp for analysis with the HumdrumToolkit (Huron, 1995).[1] It could be argued that examining an incipit is significantly different from examining a musical theme; whereas a theme is a salient, memorable musical idea, an incipit is often simply the opening melodic idea. For example, a theme dictionary, such as *Barlow and Morgenstern's Dictionary of Musical Themes* (1948), the data which many of previous nPVI studies have used, can contain the opening thematic material of multiple themes per movement. An incipit, however, which is meant to provide information for a bibliographic record, will most often contain only a single theme for a whole piece. Interestingly, it seems that, upon examining many of the themes encoded in the Barlow and Morgenstern corpus, many of the themes were cut off when the single staff of notation would run out of space. Quite interestingly, Brinkman and Huron (2018) write:

> Unfortunately, the themes tend to encode incipits rather than whole themes. The average length of the 'themes' is 19 notes, which is reasonably long compared with many other incipit-based catalogues. In browsing through the

[1] We are extremely grateful for Craig Sapp's assistance with the formatting of this data.

> Barlow and Morgenstern, our informal impression is that perhaps 5 percent of the notated themes might be consider to end at a plausible phrase boundary or cadence point. (p. 169)

In effect, these are incipits themselves, serving as record locators meant to aid the reader in finding the entire theme. It would also appear that the differences between the melodies in the Barlow and Morgenstern dataset and the RISM dataset are minimal. Whereas Brinkman and Huron found that the average length was roughly 19 notes, the average length of a RISM incipit is 14.05 noteheads, and 22.7 notes, including rests.[2] We therefore concluded that the incipits of this study are of similar length and structure to the themes found in the studies using the *Barlow and Morgenstern* dataset.

The entire RISM corpus available through this channel yielded 1,269,908 incipits – a significantly larger dataset than had been used in any of the previous studies. We excluded all duplicates (melodies that were exactly the same and listed as written by the same composer), and a separate project (Shanahan et al., in preparation) has focused on affixing extended metadata to each incipit. The entire corpus contains 16,613 composers, and each composer was checked against a 'Virtual International Authority File' online (see Bennett et al., 2006), which is meant to provide reliable information on historical figures, celebrities, *etc.* Shanahan et al. (in preparation), then wrote a web crawler that grabbed the nationality of each file, and rated the trustworthiness of the data of each file: if a human verified the data, it was given full credit, slightly less credit if two reputable online sources (such as *Oxford Online*) agreed, still less if there were mismatches, and if no data could be found it was given no credit. If we only examine the data that can be either verified by humans or has two reputable online sources agreeing, we are left with 668,350 incipits. In order to allow for future hypothesis testing with this dataset, we decided to only use a third of this dataset for the examination of our current hypotheses, analysing 225,279 randomly selected incipits. From this dataset, we only examined explicitly instrumental melodies from French, German, Italian, Czech, and English composers, written during the eighteenth-century, leaving us with 29,095 melodies for study in the current discussion.[3] This represents a dataset more than three times the size of the Barlow and Morgenstern dataset (which contains 9791 themes).

We also examined the VanHandel Art Song Corpus in conjunction with the RISM corpus (VanHandel, 2005). This corpus includes encoded art songs from 36 different French and German composers in the nineteenth century, comprising a dataset of 1164 songs.[4] This dataset has already been used to examine the effects of nPVI, and allows us to replicate these previous findings on a previously used dataset in addition to running the replication on a new corpus.

Study 1: Replicating the effect of native language on rhythmic variability

We first hoped to replicate the relationship between rhythmic variability and the primary language of a composer (as originally discussed in Patel & Daniele, 2003). For this study, we used the RISM corpus of instrumental melodic incipits, and hypothesised that the nPVI rating would be a significant predictor of composer nationality. To calculate the nPVI values, we used the same methods as Daniele and Patel (2004). At first, we limited our scope to French, German, Italian, Czech, and English composers (to anticipate our discussion a bit, we will expand this to more nationalities in the second part of Study #1). Additionally, we only used compositions written during the eighteenth century, in order to minimise as much stylistic shift as possible.

Formally stated, we hypothesised that nationality could be predicted given the nPVI rating of a melody by a French, German, Italian, Czech, or English composer. Our subsetted corpus contained 2806 French melodies; 19,620 German melodies; 3729 Italian melodies; 2529 Czech melodies, and 411 English melodies. Much like the Barlow and Morgenstern dataset (1948) there is a prevalence of German melodies. As the nPVI ratings of each nationality were not normally distributed (with each nationality providing a p-value of greater than .05 when examined with a Shapiro–Wilk test), we employed a multivariate logistic regression, which does not require an assumption of normality. With our regression model, each of the rhythmic variability ratings were used to predict one of the five nationalities.

Analysis

On the surface, the results support the hypothesis that rhythmic variability can predict composer nationality. There was a significant effect ($p < .001$; $n = 29{,}095$), but the effect size was quite small. The odds ratio for the predictors was just over 1 (1.002, with a 95% confidence interval of 1.000–1.004). We would interpret this to mean

[2] This is a count of all noteheadsand is excluding rests. It was calculated using the *census* tool from the Humdrum Toolkit (Huron, 1995).

[3] The incipit names, along with all data extracted from them, have been made publically available at: https://osf.io/fy9p3/.

[4] These corpora can be found at https://github.com/leighvh1/19th-century-art-songs-by-German-composers and https://github.com/leighvh1/19th-century-art-songs-by-French-composers. We are grateful to Leigh VanHandel for sharing this data with us.

that the odds of a composer being a certain nationality, given an nPVI rating are roughly the same (only very slightly higher) than a composer of a different nationality with a different nPVI rating. If we only examine pairs of nationalities (employing a binomial logistic regression), results are similarly significant ($p < .001$) even after correcting for multiple tests with the rather conservative Bonferroni correction, but the effect size is once again quite small. See Table 1 for a description of the odds ratio of each comparison.

Despite the large sample size, the effect size of each comparison was still quite small. Recently, scholars have been advocating for researchers to emphasise effect size over significance, as larger datasets can yield an increased likelihood of false positives in terms of significance (see Wagenmakers, 2007). In many cases, and especially cases with such a large number of observations, a *p*-value might simply be thought of as a way of seeing if we have enough statistical power to address the question at hand. We might therefore turn to methods that would illustrate the relationships of each nationality, given the rhythmic variability of the melodies.

Patel et al. (2006) argued that, in addition to rhythmic variability, the variability in pitch intervals could be a significant predictor of nationality. Measuring pitch variability through a coefficient of variation (CV), they found that nationality was better predicted when using both nPVI and CV. Put succinctly, the coefficient of variability is 'the measure of overall variability which is insensitive to the order of elements of a sequence' (Patel et al., 2006, p. 3038). With this measurement, each pitch is measured in terms of distance from a set pitch (A440), and each interval is then determined between successive onsets. The coefficient of variation is then calculated as the standard deviation of these intervals divided by the mean interval size. We therefore decided to include the coefficient of pitch variability into the analysis.

$$CV = \frac{sd(\text{int}_{\text{phrase}})}{\bar{x}(\text{int}_{\text{phrase}})} \tag{2}$$

Table 1. A list of each comparison of nationalities in pairs, with the odds ratio and the 95% confidence interval.

Comparison	Odds ratio (95% C.I)
German–French	0.9903–0.9933
German–Italian	1.0028–1.0055
German–English	0.992–0.9994
German–Czech	0.999–1.002
French–Italian	0.9943–0.9979
French–English	1.000–1.008
French–Czech	1.007–1.012
Italian–English	0.9962–1.004
Italian–Czech	1.003–1.007
English–Czech	1.0011–1.0092

Each comparison was significant at $p < .001$.

The coefficient of variation, which is measured as the standard deviation of the intervals within a phrase divided by the mean intervals of the phrase. Each interval is the distance between successive notes, each given a value as a distance from A440 in Equation (2).

When analysing VanHandel's Art Song Corpus (2005), we performed a binomial logistic regression on the entire collection of French and German art songs, using both nPVI and CV as predictors. Neither variable was significantly predictive of composer nationality in the VanHandel corpus (df = 1160, $N = 1162$, CI(nPVI) = .90–1.19; log odds ratio = 1.03); (CI(CV) = .52–1.01; log odds ratio = 0.93). This is consistent with VanHandel's original claim that the nPVI measurement was not predictive in the corpus of vocal works.

Post-hoc analyses

The effect of syllable-timed and stress-timed language

It's possible that many of the results from the analysis of the RISM corpus are obscured by the fact that, while the mean of each language might differ significantly, they contain distributions that overlap. For example, Spanish and Italian, both syllable-timed languages, might have a somewhat similar amount of rhythmic variance. As such, a multinomial logistic regression might struggle with predicting specific languages, but perhaps the broader class of 'syllable-timed/stress-timed' would be more easily predicted. As a post-hoc test, we decided to examine German, English, French, and Italian instrumental themes, the former two being stress-timed languages and the latter being syllable-timed languages.

Once again, although the means were significantly different ($p < .001$, $n = 26156$), and a logistic regression was able to predict the rhythmic group, the effect size was quite small (with an 95% CI odds ratio of 1.004–1.007). As can be seen in Figure 1 below, syllable-timed languages have a higher overall mean nPVI, which is the opposite of what we might expect to see; however, the distributions do overlap quite a bit, and there are a great deal of outliers. It would seem that the statistically significant difference is a product of the large sample size, rather than any inherent distributions.

The relationships between nationalities

Next, we decided to examine the relationships using an unsupervised clustering method. Specifically, we used hierarchical clustering to analyse the distances between average nPVI ratings. For this analysis, as we were no longer solely interested in the difference between stress-timed and syllable-timed languages, we also incorporated a number of other nationalities, including Dutch and

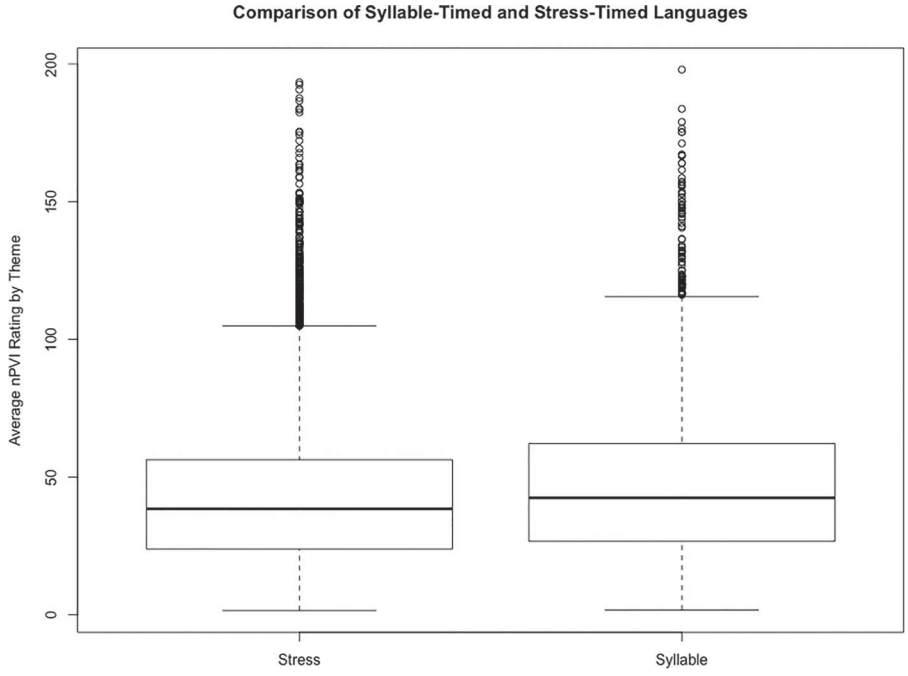

Figure 1. A comparison of nPVI ratings for syllable-timed and stress-timed languages by theme.

Swiss. The clustering of the entire century displays almost all of the relationships between languages as we would expect with the exception of German and Austrian. They are located near to one another but on separate nodes (Figure 2).

However, if we examine the two halves of the century separately, we see confounding results. In the first half of the eighteenth century we see expected relationships, such as the closeness of British and English, but Austrian and German are located far apart on separate nodes. In the second half of the eighteenth century, German and Austrian are still on separate nodes (Figure 3), and Italian becomes more differentiated in the second half of the century (Figure 4).

Change in nPVI over time

If we re-examine the studies that looked at change in rhythmic variability over time, we see a number of interesting points. Because we have so many data points, each slope is significant; however, the relationship between the slopes is a notable point of interest as well. Recall that Daniele and Patel's hypothesis was that the rising nPVI in the late eighteenth century was indicative of a move away from an Italianate style (consistent with the findings shown in Figure 4), which itself remained relatively constant. In our replication of instrumental melodies from the eighteenth century, however, we see that the Italian slope is also ascending, and is in fact doing so even more than the Austrian style. The Italian nPVI is also consistently higher than the German and Austrian values. According to their syllable-timed/stress-timed groupings, we should be seeing the opposite.

Hansen et al. (2016) had argued that a polynomial model might be a better fit. If we fit our data with such a curve, we find that Austrian and Italian follow a very similar path, with an increasing nPVI value as we head into the mid-eighteenth century. We would therefore argue that this rising nPVI that Daniele and Patel discussed could possibly be explained not as an increase, but as moving away from a dip in complexity. This dip would coincide with the Galant era, which was known for its simplicity, so perhaps we are not seeing anything regarding nationalism, but instead a more widespread change in style. We would therefore argue that nPVI seems to be more of a marker of intentional aspects of musical style, rather than linguistic origin or nationality (Figures 5 and 6).

Study 2: The effect of 'intended' national style on rhythmic variability

Musical styles were often labelled by their country of origin, such as 'French keyboard music' or 'Italian opera'. These national styles were recognised, separate styles used by composers of the time; however, these styles often transcended nationality as composers from many different parts of Europe utilised them in their compositions. Some composers even blended the national styles together to create new styles. Two composers who were

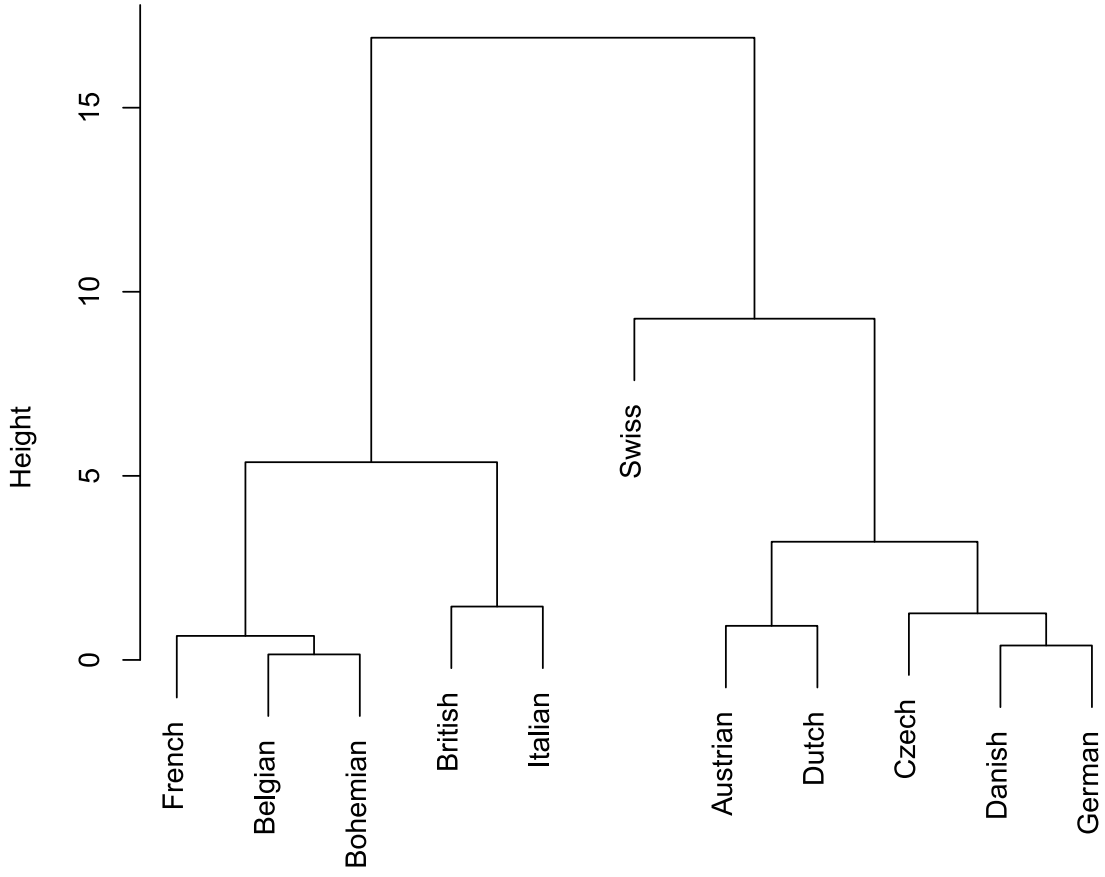

Figure 2. A hierarchical clustering of the average nPVI values of each nationality for melodies in the entire 18th century.

particularly well known for possessing this ability were Handel and Mozart.

Throughout his musical career, Handel lived and worked in many different countries, with each country having a unique impact on his compositional style (see Taruskin, 2006; chapter 5). Cudworth writes that 'Handel was one of the most eclectic of all composers, absorbing most musical influences which presented themselves to him, whether of his own or any other time, from the canzonets of Morley to the street cries of Georgian London' (Cudworth, 1959). These musical influences and combinations of different styles, including national styles are prevalent in most of Handel's music.

Like many other Viennese composers of his time, Mozart also had a reputation for blending styles. Adorno writes

> the German-Italian synthesis in his compositions has been diagnosed without pause mostly, of course, with reference to the fusion of mere categories such as opera seria, opera buffa, and Singspiel, also perhaps to the combination of southern cantability with the obligatory German way of composing, with the filigree technique of Haydn and the orchestral one of the Mannheim School. But the national moments permeate each other all the way into the smallest cells and the 'tone'. (Adorno, 1976)

Because both Mozart and Handel were able to demonstrate such flexibility across cultures, it makes sense that their music might not necessarily conform to a single 'national' rhythmic style. For this reason, it might be worth not only examining aspects of nationality, but also *intended* nationality. Our second study addresses the question of whether the intended national style would carry more weight than a composer's nationality in nPVI values. For example, if a German composer wrote a piece in the style of French keyboard music, would the nPVI values group the piece with other French keyboard music or with the music of most other German composers?

Methodology

For this study we chose to use Handel and Mozart as our two test cases. We examined Handel's English oratorios against his Italian operas and Mozart's Italian operatic style against his German operatic style. It is important to note that this study does use vocal music as opposed to instrumental music, which could present a confound in itself. But, in previous studies using vocal music such as VanHandel (2006) it has been shown that there really is no effect (it should be noted, however, that this study

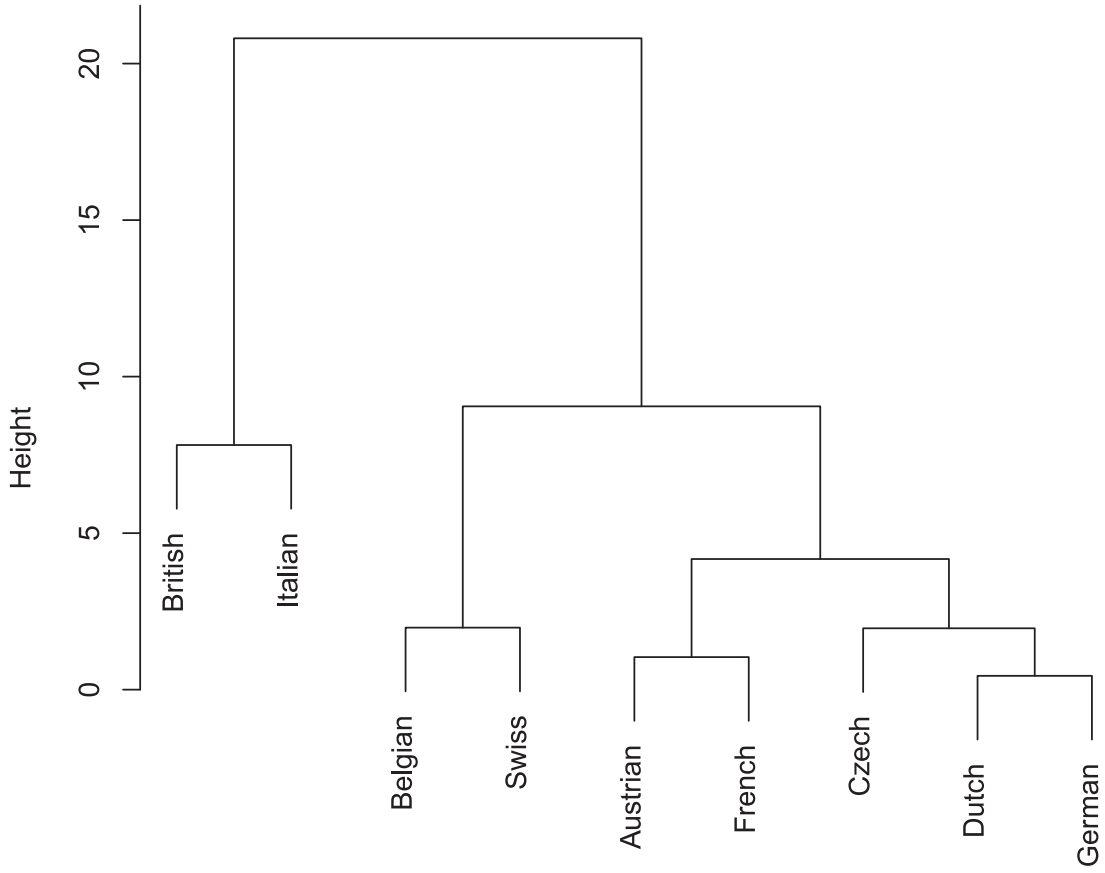

Figure 3. A hierarchical clustering of the average nPVI values of each nationality for melodies in the first half of the 18th century.

did use a different method for calculating the nPVI values). We examined 993 total vocal melodic lines from Handel's Italian and English oratorios with 488 Italian melodies and 505 English melodies. We examined 1389 vocal melodic lines from Mozart's German and Italian operas with 840 German melodies and 549 Italian melodies.

Analysis/results

We first ran a logistic regression on the Handel subset of the RISM corpus. As in study 1, results were significant ($p < .001$), but the effect size still continues to be quite small. The odds ratio for the predictors was 0.99 with the 95% CI being between .985 and 995. This would suggest that given a specific nPVI, a Handel oratorio is only slightly more likely to be labelled as a certain *intended* national style (Figure 7).

Next we ran a logistic regression on the Mozart opera set. In this case the results were not significant ($p = 0.735$), and the effect size was once again minimal. The odds ratio for the predictors was 0.99 with the 95% CI being between .995 and 1.003. These results suggest that given a specific nPVI, a Mozart opera is not significantly more likely to be labelled as a certain intended national style (Figure 8).

Conclusions

We hypothesised that nPVI would be a significant predictor of intended national style. While the results were significant with Handel's oratorios they were not with Mozart's operas, although the effect size was very small in both cases. Our results were fairly inconclusive, but we have derived several theories as explanation. One such theory is that we could be seeing a gradual shift away from specifically national styles in the post-Baroque era. Although Handel was fluent in both German and English, his Baroque compositions are situated within a time in which nationality commonly served as a defining characteristic of style. However, after this period, instead of strictly adhering to their own national styles, the composers of the Classical era, including Mozart, began to utilise and blend multiple styles. Their compositional identities became influenced less by country and more of a result of personal stylistic choices and preferences. This then lead to the development of schools of composition characteristic of the Romantic period.

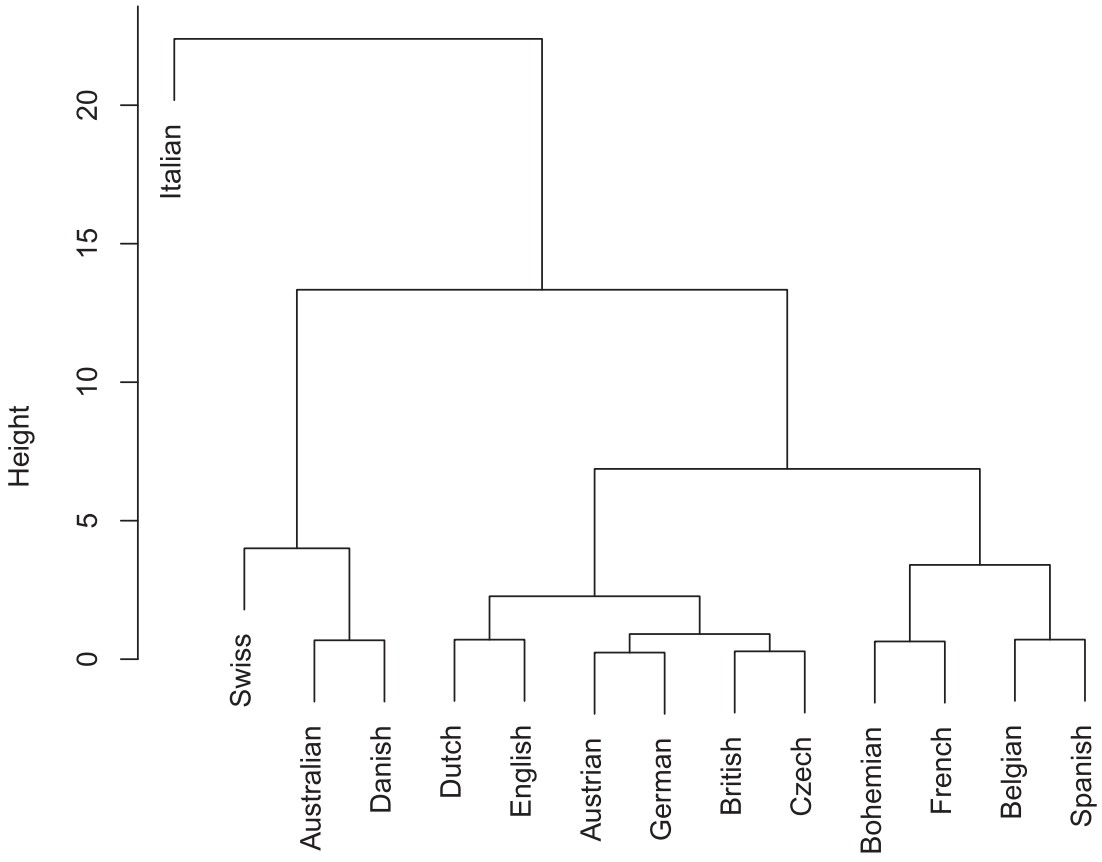

Figure 4. A hierarchical clustering of the average nPVI values of each nationality for melodies in the second half of the 18th century (note how Italian becomes more differentiated).

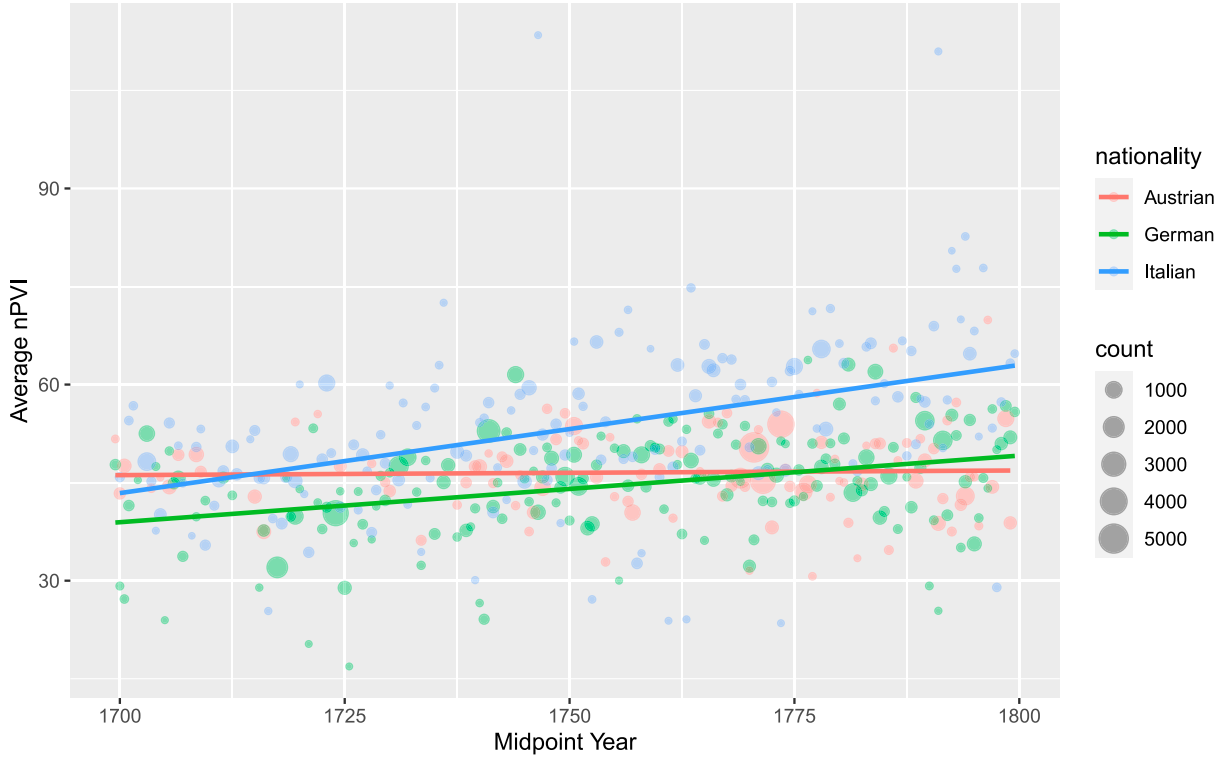

Figure 5. A linear fit of 18th century French, German and Italian nPVI values over time.

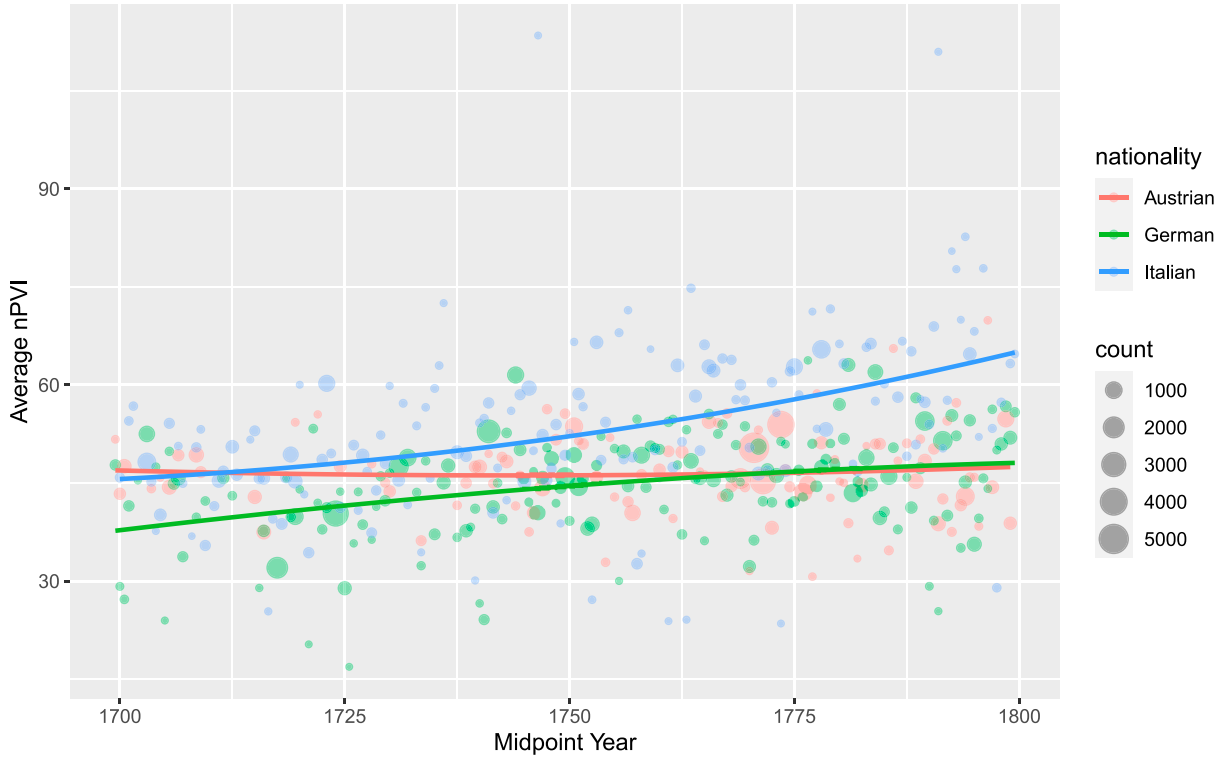

Figure 6. A moving average fit of 18th century French, German and Italian nPVI values over time.

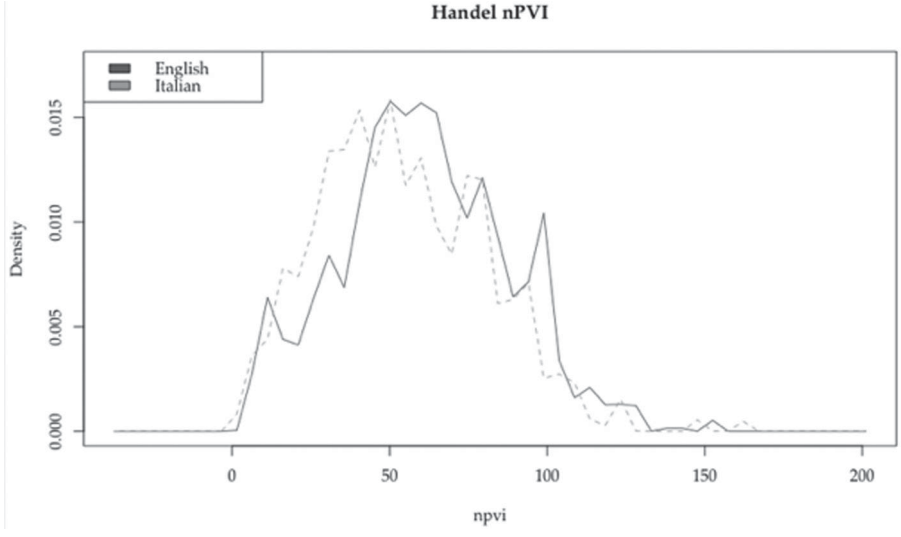

Figure 7. A density plot of the nPVI values of Handel's English and Italian oratorios.

Composers no longer identified by country alone, but also by their chosen style and school of composition, such as the Brahmsian and Wagnerian schools in what is often called the War of the Romantics (see VanHandel, 2017). Another possibility is that these results are due to other stylistic features at play. The nPVI metric only accounts for rhythmic variance, and most styles have many additional distinguishing features besides rhythm alone.

Study 3: The addition of pitch-transition probabilities

In order to replicate this, we resampled the RISM dataset taking 25,000 short incipits. Excluding all melodies without clearlynotated nationalities,we were left with 21,572 melodies, including 10,877 German melodies, 1114 French, 6207 Italian, 1225 Czech, and 2149 English melodies.

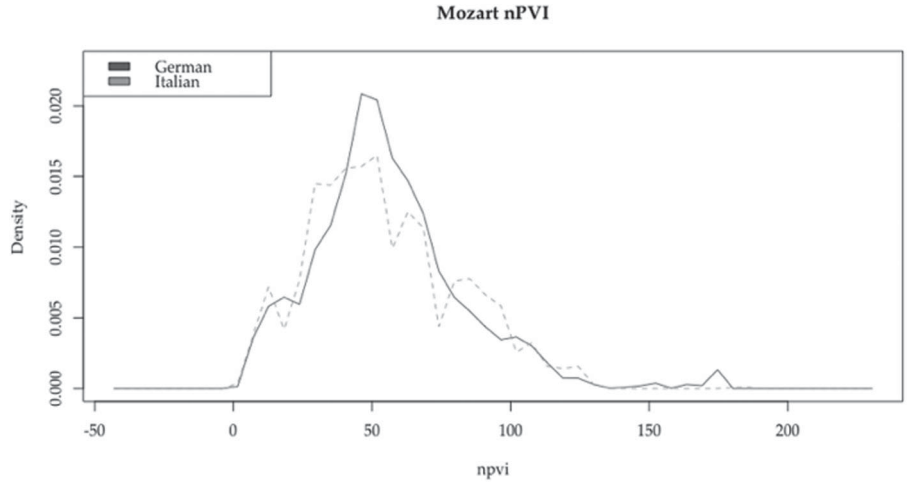

Figure 8. A density plot of the nPVI values of Mozart's German and Italian operas.

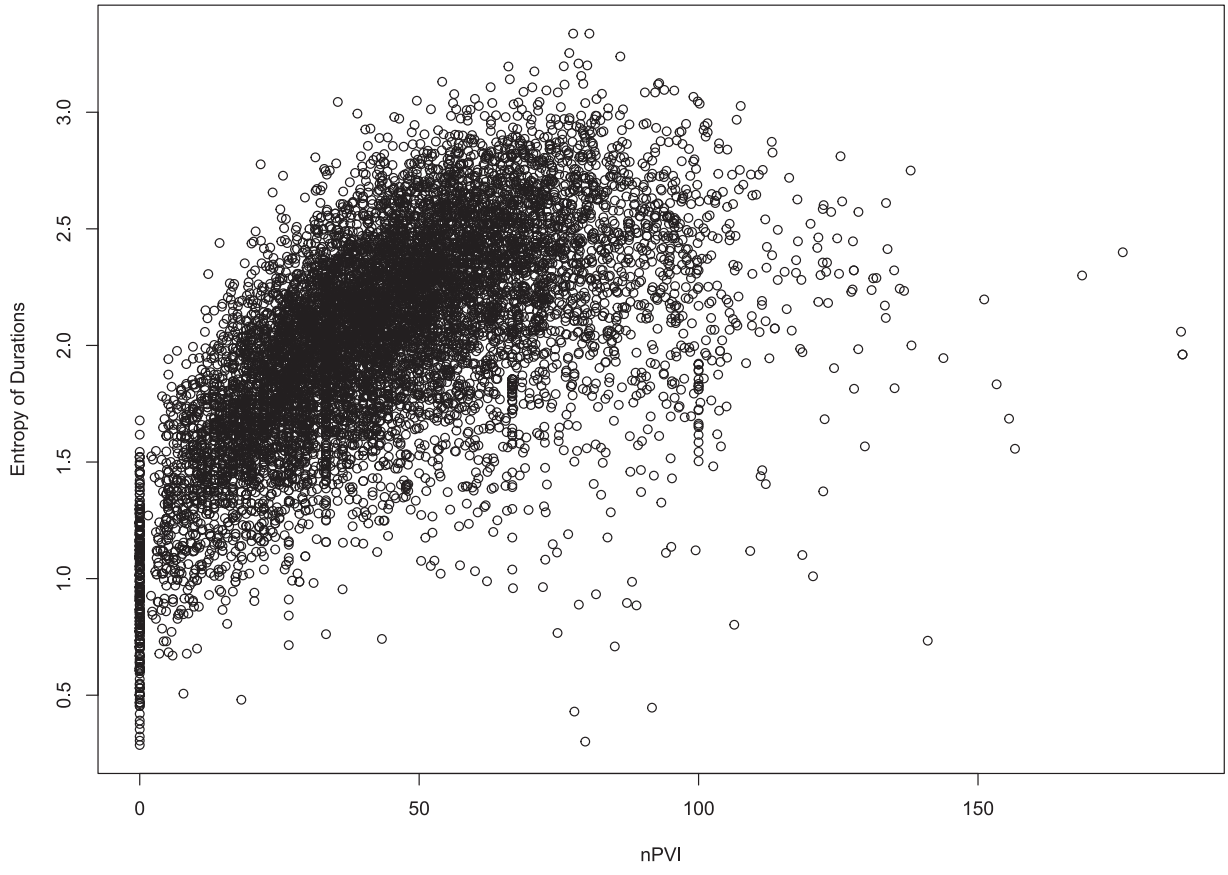

Figure 9. Correlation between nPVI and entropy of rhythmic durations in a random sample of 10,000 incipits.

The coefficient of variability was calculated as the standard deviation of the pitch intervals divided by the mean value of all the pitches in a single melodic incipit. In order to quantify the melodies, each pitch was first given a value by calculating its interval from a specific pitch. Although Patel et al. (2006) used A440, we used middle C (261.6 Hz) as our reference pitch (while it's important that the variability is calculated around a single pitch, the value of that pitch is not important, as long as the relative positions of the tones within the melodic incipit are maintained). Using these values, we calculated the interval between adjacent note pairs, then calculating the standard deviation between the intervals. This standard deviation was then divided by the mean to give us our coefficient of variability.

Results

Patel et al. (2006, p. 3039) performed a linear regression to examine how well linguistic origin might be predicted by both nPVI and melodic variability. We replicated these results using a logistic regression in which both melodic and rhythmic variability predicted linguistic origin, and found that, while results were significant, the effect size was again quite small ($p < .001$; $R^2 = .009$; df = 21,570). The log odds ratio for the predictors was less than 1 at 0.953 with a 95% confidence interval of 0.901–1.010. We would interpret these results to mean that using the coefficient of variability in addition to the nPVI values does not markedly affect the accuracy of predictions of nationality. It would seem that, through the analysis of short incipits, rhythmic and intervallic variability might be somewhat predictive of nationalistic and linguistic origins, but the effect size seems to be quite small.

The relationship of nPVI and complexity

Whether or not there is a relationship between national origin, the nPVI measurement does seem to provide a measurement of rhythmic complexity, which can be seen as changing over time and through different social circles (see VanHandel, 2017). We might therefore examine the relationship between another measurement of complexity and the nPVI measurement. Shannon Entropy has been used as a broad measurement of complexity of music since the early 1950s (see Hiller & Bean, 1966; Youngblood, 1958), and it would make sense that rhythmic entropy would strongly correlate with nPVI of a theme. To explore this further, we took a random subsample of 10,000 incipits, and looked at the correlation between nPVI and the entropy of rhythmic durations.[5] There does seem to be a strong correlation between the two ($R^2 = .40$; $p < .001$; df = 9998). See Figure 9 below.

When looking at the VanHandel Art Song Corpus, however, we found that the correlation between the durational entropy (in which the Shannon entropy is calculated of successive durations) is not significantly correlated ($N = 1164$, df = 1162, $p = .11$, $R^2 = 0.001$). This is likely due to the fact that the Art Song corpus contains complete works, and there is significant variation in durations and rhythmic variability throughout an entire piece. It has been proposed that global measurements of entropy are less effective than models that incorporate a 'viewpoint' model (Conklin & Witten, 1995), and it is likely that this is the case with durational entropy and rhythmic variance, as well. When looking at smaller incipits in the RISM collection, however, the significant correlation between durational complexity and the nPVI rating suggests that nPVI would be a complementary method for analysing rhythmic complexity in a way that overlaps – although not entirely – with information-theoretic metrics for complexity.

[5] The data for both of these calculations can be found on the Open Science Framework page for this article at https://osf.io/fy9p3/.

Discussion

The validity of the nPVI metric has been questioned with increasing frequency over the past few years, and the results of the studies of this paper seem to be in alignment with much of the recent work showing that a relationship between linguistic and melodic variability is not as straightforward as one would like. Similar to other metrics of variability, such as Shannon Entropy (Shannon, 1951), the uses of the metric have gone beyond simply music and language comparisons. Shannon Entropy was originally intended to compress communication signals into transmittable, but meaningful sets of information. Its use was to facilitate communication; nevertheless, it has been used to measure musical complexity (Hiller & Bean, 1966; Youngblood, 1958), as well as literary complexity and visual complexity (Arnheim, 1974).

In the theory and philosophy of measurement, there is a concept known as the 'representational theory of measurement' (Tal, 2017), which suggests that the origins of a measurement are not always necessary. For example, it doesn't matter why a metre or a foot are the lengths they are, it only matters that those units represent a consistent structure. Although linguistic variability might be able to be addressed with the nPVI, it seems unclear how we might measure the national origins of short musical themes with such a measure, and the recent reassessments of the theme seems to suggest this as well (Condit-Schultz, 2016, 2019; Temperley, 2017). Nevertheless, the nPVI provides a framework in which we might examine the changing relationships of complexity (over time, by instrument, by genre, *etc.*). Like entropy, the usefulness of the measurement might exceed the original intent.

Notes

1. We are extremely grateful for Craig Sapp's assistance with the formatting of this data.
2. This is a count of all noteheadsand is excluding rests. It was calculated using the *census* tool from the Humdrum Toolkit (Huron, 1995).
3. The incipit names, along with all data extracted from them, have been made publically available at: https://osf.io/fy9p3/.
4. These corpora can be found at https://github.com/leighvh1/19th-century-art-songs-by-German-composers and https://github.com/leighvh1/19th-century-art-songs-by-French-

composers. We are grateful to Leigh VanHandel for sharing this data with us.

5. The data for both of these calculations can be found on the Open Science Framework page for this article at https://osf.io/fy9p3/.

Acknowledgments

The authors wish to thank Craig Sapp for his valuable assistance with the data cleaning, Leigh VanHandel for providing us with the art song data, Ani Patel for assistance with clarifying the CV metric, and two anonymous reviewers for their helpful feedback.

References

Adorno, T. W. (1976). *Introduction to the sociology of music*. The Seabury Press.

Arnheim, R. (1974). *Entropy and art: An essay on disorder and order*. University of California Press.

Arvaniti, A. (2012). The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics*, *40*(3), 351–373. https://doi.org/10.1016/j.wocn.2012.02.003

Barlow, H., & Morgenstern, S. (1948). *A dictionary of musical themes*. Crown.

Bennett, R., Hengel-Dittrich, C., O'Neill, E. T., & Tillett, B. B. (2006). Viaf (virtual international authority file): Linking die deutsche bibliothek and library of congress name authority files. World Library and Information Congress: 72nd IFLA General Conference and Council.

Brinkman, A., & Huron, D. (2018). The leading sixth scale degree: A test of Day-O'Connell's theory. *Journal of New Music Research*, *47*(2), 166–175.

Brook, B. S. (1965). The simplified "Plaine and Easie code system" for notating music: A proposal for international adoption. *Fontes ArtisMusicae*, *12*(2/3), 156–160.

Condit-Schultz, N. (2016). Deconstructing nPVI. *Proceedings of the 14th International Conference on Music Perception and Cognition* (pp. 800–804). ICMPC.

Condit-Schultz, N. (2019). Deconstructing the nPVI: A methodological critique of the normalized pairwise variability index as applied to music. *Music Perception: An Interdisciplinary Journal*, *36*(3), 300–313. https://doi.org/10.1525/mp.2019.36.3.300

Conklin, D., & Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, *24*(1), 51–73. https://doi.org/10.1080/09298219508570672

Cudworth, C. (1959). Handel and the French style. *Music & Letters*, *40*(2), 122–131. https://doi.org/10.1093/ml/XL.4.122

Daniele, J. R. (2017a). A tool for the quantitative anthropology of music: Use of the nPVI equation to analyze rhythmic variability within long-term historical patterns in music. *Empirical Musicology Review: EMR*, *11*(2), 228. https://doi.org/10.18061/emr.v11i2.4893

Daniele, J. R. (2017b). The "rhythmic fingerprint": An extension of the nPVI to quantify rhythmic influence. *Empirical Musicology Review: EMR*, *11*(2), 243. https://doi.org/10.18061/emr.v11i2.5426

Daniele, J. R., & Patel, A. D. (2004). The interplay of linguistic and historical influences on musical rhythm in different cultures. *Proceedings of the 8th International Conference on Music Perception and Cognition* (pp. 759–762).

Daniele, J. R., & Patel, A. D. (2013). An empirical study of historical patterns in musical rhythm. *Music Perception: An Interdisciplinary Journal*, *31*(1), 10–18. https://doi.org/10.1525/mp.2013.31.1.10

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In *Laboratory Phonology* (pp. 515–546).

Hansen, N. C., Sadakata, M., & Pearce, M. (2016). Nonlinear changes in the rhythm of European art music: Quantitative support for historical musicology. *Music Perception: An Interdisciplinary Journal*, *33*(4), 414–431. https://doi.org/10.1525/mp.2016.33.4.414

Hiller, L., & Bean, C. (1966). Information theory analyses of four sonata expositions. *Journal of Music Theory*, *10*(1), 96–137. https://doi.org/10.2307/843300.

Huron, D. (1995). *The humdrum toolkit: Reference manual*. Center for Computer Assisted Research in the Humanities.

Huron, D. (2013). On the virtuous and the vexatious in an age of big data. *Music Perception: An Interdisciplinary Journal*, *31*(1), 4–9. https://doi.org/10.1525/mp.2013.31.1.4

Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele (2003). *Music Perception: An Interdisciplinary Journal*, *21*(2), 267–271. https://doi.org/10.1525/mp.2003.21.2.267

London, J. (2013). Building a representative corpus of classical music. *Music Perception: An Interdisciplinary Journal*, *31*(1), 68–90. https://doi.org/10.1525/mp.2013.31.1.68

London, J., & Jones, K. (2011). Rhythmic refinements to the nPVI measure: A reanalysis of Patel & Daniele (2003a). *Music Perception: An Interdisciplinary Journal*, *29*(1), 115–120. https://doi.org/10.1525/mp.2011.29.1.115

Patel, A. D. (2008). *Music, language, and the brain*. Oxford University Press.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, *87*(1), B35–B45. https://doi.org/10.1016/S0

Patel, A. D., Iversen, J. R., & Rosenberg, J. C. (2006). Comparing the rhythm and melody of speech and music: The case of British English and French. *The Journal of the Acoustical Society of America*, *119*(5 Pt 1), 3034–3047. https://doi.org/10.1121/1.2179657

Répertoire International des Sources Musicales. (2018). Retrieved May, 2015 http://www.rism.info/en/home.html

Shanahan, D., Sapp, C., & Bell, E. (in preparation). Curating the RISM dataset for questions of stylistic change.

Shanahan, D., & Shanahan, E. (2014). The densmore collection of native American songs: A new corpus for studies of effects of geography and social function in music. *Proceedings for the 13th International Conference for Music Perception and Cognition* (pp. 206–209).

Shannon, C. E. (1951). Prediction and entropy of printed English. *Bell System Technical Journal*, *30*(1), 50–64.

Tal, E. (2017). Measurement in science. In E. N. Zalta (Ed.), *The stanford encyclopedia of philosophy*. https://plato.stanford.edu/archives/fall2017/entries/measurement-science/

Taruskin, R. (2006). *Music in the seventeenth and eighteenth centuries: The oxford history of western music*. Oxford University Press.

Temperley, D. (2017). Rhythmic variability in European vocal music. *Music Perception: An Interdisciplinary Journal*, *35*(2), 193–199. https://doi.org/10.1525/mp.2017.35.2.193

Temperley, N., & Temperley, D. (2011). Music-language correlations and the "scotch snap.". *Music Perception: An Interdisciplinary Journal*, *29*(1), 51–63. https://doi.org/10.1525/mp.2011.29.1.51

Toussaint, G. T. (2012). The pairwise variability index as a tool in musical rhythm analysis. *Proceedings of the 12th International Conference on Music Perception and Cognition* (pp. 1001–1008).

VanHandel, L. (2005). *Setting a menu to music: Prosody and melody in 19th Century art songs* [PhD Dissertation]. Stanford University.

VanHandel, L. (2006). *Trends in/over time: Rhythm in speech and music in 19th century art song*. Proceedings of the 9th International conference on music perception and cognition. Bononia University Press.

VanHandel, L. (2017). The war of the romantics: An alternate hypothesis using nPVI for the quantitative anthropology of music. *Empirical Musicology Review: EMR*, *11*(2), 234. https://doi.org/10.18061/emr.v11i2.5474

VanHandel, L., & Song, T. (2010). The role of meter in compositional style in 19th century French and German art song. *Journal of New Music Research*, *39*(1), 1–11. https://doi.org/10.1080/09298211003642498

Wagenmakers, E.-J. (2007). A practical solution to the pervasive problems of *p* values. *Psychonomic Bulletin & Review*, *14*(5), 779–804. https://doi.org/10.3758/BF03194105

Youngblood, J. E. (1958). Style as information. *Journal of Mathematics & Music. Mathematical and Computational Approaches to Music Theory, Analysis, Composition and Performance*, *2*(1), 24–35. https://doi.org/10.2307/842928.