Rhythmic Refinements to the nPVI Measure: A Reanalysis of Patel & Daniele (2003a)

Justin London & Katherine Jones
Carleton College

The normalized pairwise variability index (nPVI), originally developed for the analysis of speech prosody, has been shown to be a useful tool in the analysis of the rhythmic variability present in individual melodies and melodic corpuses (Patel & Daniele, 2003a; Huron & Ollen, 2003; Patel, Iverson, & Rosenberg, 2006). Here we suggest a number of rhythmic refinements for the application of the nPVI in musical contexts. These include (a) applying the nPVI metric to higher levels of rhythmic structure, (b) taking metrical structure (duple versus triple) into account, and (c) using alternative codings for surface durations. A reanalysis of the data from Patel & Daniele (2003a) is given as a demonstration of the utility of these refinements for the analysis of musical styles and repertoires; other possible refinements are discussed.

Received September 8, 2010, accepted April 28, 2011.

Key words: meter, rhythm, variability, nPVI, hierarchy

The normalized pairwise variability index (nPVI,) was originally developed to measure the variability of surface durations in language (Low, Grabe, & Nolan, 2000; Grabe & Low, 2002), a more nuanced approach than the stress- versus syllable-timed distinction that is often made between various languages and dialects (Roach, 1982). Grabe and Low (2002) used the nPVI to compare successive vocalic intervals measured from the onset of a vowel to its offset, whether it included a monophthong, diphthong, or even greater number of vowels. Their nPVI-based analyses revealed a distinction between stress-timed versus syllable-timed languages, with a higher nPVI for syllable-timed languages than stress-timed languages. Moreover, Grabe and Low's findings support a continuum wherein a language could be more or less stress-timed, as opposed to a dichotomous split between stress- versus syllable-timed languages.

The calculation of the nPVI value for an utterance is straightforward. As the aim is to capture the durational variability present in a sample of continuous speech, the nPVI looks at the difference in duration between each successive pair of speech units. To do so one could simply look at the absolute value of each difference: $|D_n - D_{(n+1)}|$ for each pair, and then calculate the average durational difference for the entire utterance. Two problems emerge, however. First, using actual durations makes it difficult (if not impossible) to compare variability for speakers with different baseline speech rates. Second, speakers often accelerate their speech rate over the course of a breath group, distorting the variability. Thus the "n" of the nPVI involves normalizing each durational difference by dividing the absolute value of each difference by the average duration of each pair of durations: $|D_n - D_{(n+1)}| \div [(D_n + D_{(n+1)}) \div 2]$. As one can see, if the two durations are identical, the value is 0; if the ratio between the two is 1:3 the normalized value is 1, and as the shorter value approaches zero (or the longer value approaches infinity), the maximum normalized difference is 2. Again, for an entire speech sequence the normalized differences are then averaged, and by convention, multiplied by 100. Grand averages then can be calculated for a corpus of utterances and the characteristic nPVI for a speaker, dialect, or language may be obtained. As Patel and Daniele (2003a) note, "this normalization … makes the nPVI a dimensionless quality which can be applied to both language and music" (p. B37).

In a number of articles, Aniruddh Patel and his colleagues have done just that (Patel, 2003; Patel & Daniele, 2003a, 2003b; Patel, Iverson, & Rosenberg, 2006). They have shown both (a) that there are differences in the variability of surface rhythm in late 19th and early 20th century music by French versus British composers, and (b) that these differences can be correlated with the difference in nPVI between French and (British) English more generally. Rather than using durational values from recorded speech, they used the rhythmic values indicated in musical scores, as scores give precise ratios for their relative duration (a potential pitfall of

Music Perception volume 29, issue 1, pp. 115–120. issn 0730-7829, electronic issn 1533-8312. DOI: 10.1525/mp.2011.29.1.115

this methodology is noted below, though we have also followed it here). Patel and his colleagues based their studies on a sample of 181 French and 137 English melodies taken from the Barlow and Morgenstern *Dictionary of Musical Themes* (1948). Huron and Ollen (2003) expanded the corpus to almost 2000 melodies and found similar differences in French versus English musical nPVI, while Hannon (2009) showed that the differences in surface rhythms were perceptually salient in a test using novel examples of French versus English music as stimuli.

Clearly, the nPVI is a useful way of analyzing the surface variability in a variety of rhythmic domains (whether speech, language, and perhaps other modalities, such as visual rhythms). But there are a number of crucial differences in the rhythmic structure of music versus linguistic prosody:

- The timings of surface events in music are usually linked to one or more isochronous periodicities, while speech events are not;
- Rhythmic and metrical organization in music is highly constrained, in most instances based on a binary or ternary organization of beats, beat subdivisions, and measures (London, 2001, 2004).

The first point is illustrated in Figure 1. This is an example based on Patel and Daniele (2003a) that shows how the durational values from the notated score were coded for use with the nPVI formula. The top line of figures (below the staff) gives the surface (S) durations, indexed to the quarter note (quarter note = 1). The second line gives our codings for the durations on the next highest level of rhythmic structure (S+1), based on our analysis of its rhythmic grouping structure (Cooper & Meyer, 1960; Lerdahl & Jackendoff, 1983); notice how these reflect the isochronous tactus. Indeed, in all metrical music as one moves up the rhythmic hierarchy, the average nPVI value for a melody typically regresses to zero as all group durations become equivalent to the length of the measure.

The nPVI formula may be used on higher levels of rhythmic structure, and the analysis of nPVI data may include a consideration of metrical type (duple versus triple); this is what we have done in the current study. Using the same corpus as Patel & Daniele (2003a), the rhythmic structure of each melody was checked by examining the incipits published in the Barlow and Morgenstern dictionary (1948). A slightly smaller corpus was used in this study, as a number of melodies in irregular meters (e.g., 7/4, 5/4) or with other coding problems were excluded. The remaining 286 melodies (116 British, 170 French) were then classified according to their metrical types as either duple or triple (number of beats per measure) and as either simple or compound (duplet versus triplet beat subdivision), as indicated in Table 1. For most melodies the determination of metrical type was straightforward, based on their use of standard time signatures and normative rhythmic figures. However, a few time signatures are ambiguous with respect to metrical type (Mirka, 2009). At slow tempos, 3/8 may represent simple triple meter, but at faster tempos pairs of 3/8 measures may effectively project a sense of 6/8; this was not the case with any examples in our data set. The 6/4 time signature gives rise to the opposite problem: at rapid tempos it may be analogous to 6/8, but at slower tempos it may be heard as 3/4 + 3/4. Based on rhythmic analyses performed by the first author (a music theorist), melodies in 6/4 were sorted into simple triple versus compound duple time, taking their characteristic durational patternings into account; this can be seen in Table 1.

A higher level grouping analyses for each melody was performed, categorizing its rhythms in terms of basic rhythmic archetypes (Cooper & Meyer, 1960), or in cases where surface durations were continuously isochronous, extracting higher level periodicities based on the prevailing metrical type. From these analyses S+1 duration series were obtained. In most instances the S+1 level was either the tactus (for melodies with one or more levels of beat subdivision) or the half measure; rarely was S+1 equal to the notated measure. It is also worth noting that for 29% of all melodies (30% British, 28% French) the S+1 nPVI was zero—in other words, almost a third of the higher level group durations were isochronous. By

FIGURE 1. Barlow and Morgenstern (1948) Example D122, Debussy's String Quartet in G-minor, first movement/2nd theme, after Patel and Daniele (2003a) Figure 2, p. B40.

TABLE 1. Metrical distribution of melodies used in the current nPVI analyses.

	Duple		Triple	
Subdivision	Signature	Count	Signature	Count
Simple	4/4	95	3/4	55
	2/4	40	3/8	9
	2/2	27	3/2	7
	4/8	4	6/4	2
	2/8	1		
	Total	**167**	Total	**73**
Compound	6/8	28	9/8	7
	6/4	6	9/16	1
	12/8	4		
	Total	**38**	Total	**8**

contrast, only 5% of the surface nPVI values were zero; very few melodies had wholly isochronous surface rhythms.

A one-way ANOVA of our (slightly) reduced data set confirmed Patel and Daniele's (2003a) main result for surface level durations: mean British nPVI = 46.97 (n = 116) vs. mean French nPVI = 39.92 (n = 170): $F(1, 284) = 5.94$, $p = .015$. As one might expect, at the S+1 level there was no significant difference between British and French rhythmic variability (British nPVI = 22.85 vs. French nVPI = 21.31; $F(1, 284) = 0.36$, $p = .548$). A one-way ANOVA of the full data set showed no main effect for metrical type (duple versus triple) on the surface (S) level (mean duple nPVI = 41.35, n = 206; mean triple nPVI = 46.47, n = 80; $F(1, 284) = 2.59$, $p = .109$). However, there was a strong main effect for metrical type at the S+1 level (mean duple nPVI = 18.56 vs. mean triple nPVI = 30.65; $F(1, 284) = 20.47$, $p < .001$). This makes sense, as the triple measures afford a higher-level long-short grouping structure that duple measures do not. Note also that the mean S+1 nPVI values are lower than the S level nPVI averages in both duple and triple contexts; these differences were highly significant ($p < .001$) in both.

One can probe a bit deeper, and look at the effect of metrical type within each nationality; see Figures 2a and 2b. In the English melodies there are no significant differences between duple and triple melodies on either the S or S+1 levels; in part this may be due to the limited sample size of British triple melodies (n = 26). However, in the French melodies the differences between duple (n = 116) and triple (n = 54) are significant both on the S and S+ 1 levels (one-way ANOVA, $F(1, 168) = 4.37$,

A

B

C

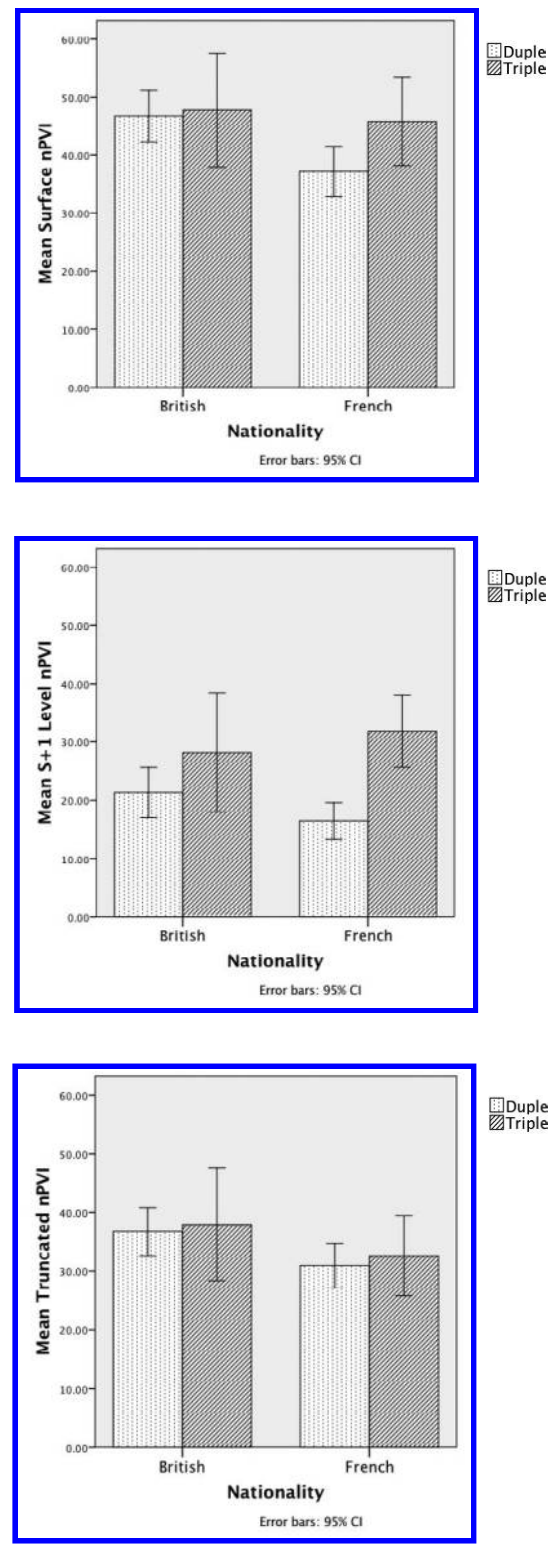

FIGURE 2. nPVI comparisons by nationality (British versus French). A) Surface nPVI comparisons. B) S+1 level nPVI comparisons. C) Truncated surface nPVI comparisons.

p = .038 and $F(1, 168) = 23.87$, $p < .001$, respectively). As there was some concern regarding the non-normal distribution of the data, an independent samples Mann-Whitney U test confirmed the significance of these results at the S (Z = -2.04, p = .042) and S+1 (Z = -4.18, p < .001) levels.

Reversing the dependent and independent factors, results of a one-way ANOVA between French (n = 54) versus English (n = 26) melodies in triple meters were nonsignificant on both S and S+1 levels, $F(1, 78) = 0.09$, p = .763 for S and $F(1, 78) = 0.43$, p = .517 for S+1; see Figures 3a and 3b. However, the nPVI differences for duple meters (British: n = 90; French: n = 116) were highly significant on the S level, $F(1, 204) = 9.06$, p = .003, and nearly significant on the S+1 level, $F(1, 204)$ = 3.39, p = .067. The low number of English melodies in triple meters precluded running any two-way or three-way ANOVAs to test for significant interactions; this is a reflection of the relatively lower number of triple meter pieces in music more generally. Nonetheless, given the robustness of results of the one-way ANOVAs on the subsets of the data, we would argue that the difference between the French and English melodies originally reported by Patel and Daniele (2003a), and confirmed here, is due to the nPVI differences in the duple, and not the triple melodies in the corpus.

In addition to taking metrical type and metrical/rhythmic level into account, the surface rhythms may be coded in other ways that reflect the influence of the beat or tactus levels and gives consideration to group boundaries. Figure 4 illustrates how this may be done. The "raw" nPVI coding, analogous to that used used by Patel and his colleagues, simply gives interonset intervals between every note in the melody. However, as this melody is comprised of four distinct subgroups with gaps between them, the "group" nPVI coding truncates the group-final duration, rounding each off to the nearest beat. By making the terminal duration of each group equal to the length of the tactus, we avoid giving group-final durations undue weight, and we avoid conflating within versus between group durational intervals. This truncation gives rise to dramatically different nPVI averages for the melody, as the raw values give an nPVI of 96.52, while the group values give a value of 52.06. Likewise, to the extent to which a continuous pulse is felt over the course of the passage, an even greater sense of rhythmic regularity may be felt by a listener than is indicated by this lower nPVI value.

In sorting the corpus into the metric categories as noted above, we were also able to recode all surface durations relative to the tactus level, assigning tactus durations a value of 1. This did not affect the calculation of the S and S+1 nPVI values, given its durational normalization.

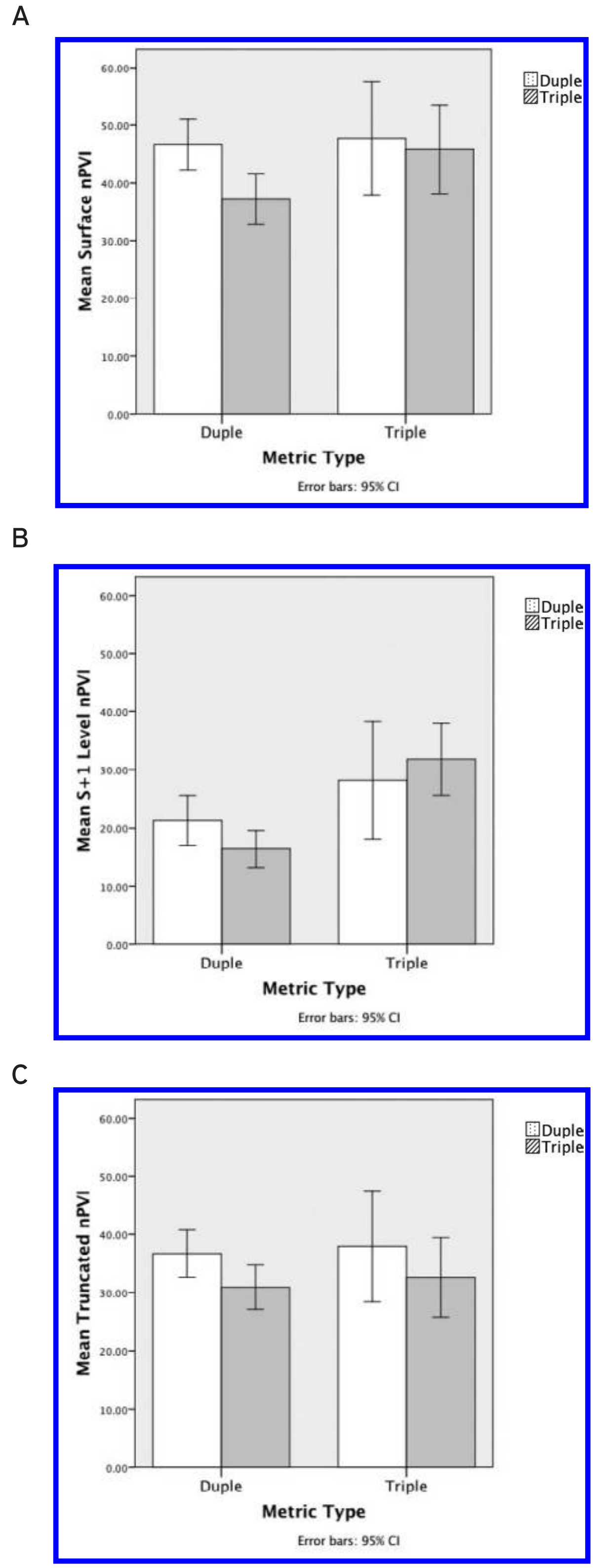

FIGURE 3. nPVI comparisons by metrical type (duple versus triple). A) Surface nPVI comparisons. B) S+1 level nPVI comparisons. C) Truncated surface nPVI comparisons.

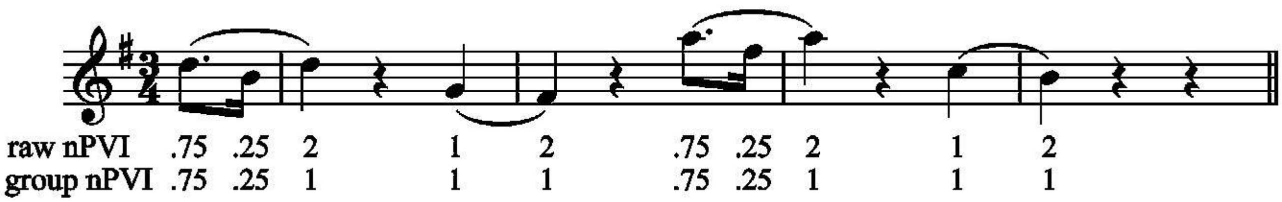

FIGURE 4. Alternative codings for the surface rhythms, Mozart's Piano Sonata in G-major, K. 283, first theme. First line of figures reflects "raw" durational (inter-onset interval) values; second line truncates durations to within-group durations, rounded to the nearest beat.

However, this recoding provided a ready means of reanalyzing the corpus by simply truncating any durations greater than 1 (i.e., giving them a value of 1) and then recalculating the nPVI value for each melody. In the "truncated" data set, a main effect for nationality was still present, albeit somewhat reduced: $F(1, 284) = 4.61$, $p = .033$; again, there was no significant effect for metrical type. Within the British subset the difference between duple and triple meters remained nonsignificant, and within the French subset the difference between duple and triple meters was no longer significant; compare Figures 2a and 2c. Within the duple meter subset the difference between French and British melodies remained significant, $F(1, 204) = 4.20$, $p = .042$. Within the triple meter subset there was no significant difference between French and British melodies; compare Figures 3a and 3c. This is informative, as it implies that the main effects previously observed on the S level are due to durational variations within rhythmic groups, and not due to variations in the intervals between rhythmic groups.

The nPVI measure may be usefully expanded/enhanced in its analysis of musical rhythm, beyond its applicability to the study of music-language relationships. As such, the nPVI metric may be a useful tool for music analysis in its own right (e.g., VanHandel, 2009). Here we have shown that in applying nPVI to music, differences between surface and higher levels of rhythmic structure can and should be taken into account, as well as differences between metrical types. In the case of Patel and Daniele's (2003a) data set, our further analysis shows that the difference between the British versus French melodies is due to differences in duple as opposed to triple meters. The lack of difference between duple and triple British melodies is also surprising, as one would expect triple melodies to be inherently more "uneven," especially at the S+1 level; this suggests a topic for further musical analysis and scrutiny.

Other areas for future research include the use of performance timing information in calculating nPVI values, rather than score-based durations. As is well known, in live performance, notated durations are altered via expressive variations in timing and dynamics. While successive durations may be notated as equivalent, in live performance one never gets zero durational differences (and the concomitant overall reduction in nPVI value). Indeed, in many styles, from swung rhythms in jazz to the characteristic rubato in Chopin's piano music, deviations from notated isochrony may be considerable. Whether using score-based or performance-based input, the nPVI metric may be useful in developing algorithms that can encode both surface and higher-level rhythmic durations, an important goal of computer-based music analysis. Finally, one may wish to explore alternatives to the nPVI formula, especially if one is interested in tempo-sensitive measures of durational variability. As noted above, the "n" of the nPVI formula refers to the normalization of successive durations, achieved by dividing the absolute value of the difference of each duration by their average length. This also means that any integer value may be used as durational value—one may index durations to the quarter note, and use fractions for smaller notes, or use the smallest common duration as "1" and then employ multiples for all higher values; in either case the nPVI values will be the same as a result of the normalization. Yet there are important differences in the perception and cognition of durations on different time scales and at different tempos (see London, 2002, 2004). Normalization will also distort the effect of accelerandi, ritardandi, and expressive timing nuances, as a gradual change in duration results in a very small nPVI, though the end result may be a dramatic shift in durational pacing. Thus there may be some value in "de-normalizing" the PVI formula to capture distinctions of temporal scale and pacing that are musically salient. Suffice to say that researchers have long known that musical rhythm is subtle and complex. By extending the nPVI metric, we may have a powerful tool with which to capture many of these subtleties and nuances.

Author Note

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

References

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

Cooper, G., & Meyer, L. B. (1960). *The rhythmic structure of music.* Chicago, IL: University of Chicago Press.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the rhythm class hypothesis. In C. Gussenhoven & N. Warner (Eds.), *Laboratory phonology* (pp. 515–546). Berlin: Mouton de Gruyter.

Hannon, E. E. (2009). Perceiving speech rhythm in music: Listeners classify instrumental songs according to language of origin. *Cognition, 111*, 403–409.

Huron, D., & Ollen, J. (2003). Agogic contrast in French and English themes: Further support for Patel and Daniele. *Music Perception, 21*, 267–271.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

London, J. (2001). Rhythm. In J. Tyrrell & S. Sadie (Eds.), *The new Grove dictionary of music and musicians* (Vol. 21, 2nd ed., pp. 277–309). Oxford, UK: Oxford Music University Press.

London, J. (2002). Cognitive constraints on metric systems: Some observations and hypotheses. *Music Perception, 19,* 529–550.

London, J. (2004). *Hearing in time.* New York: Oxford University Press.

Low, E. L., Grabe, E., & Nolan, F. (2000). Quantitative characterisations of speech rhythm: 'Syllabletiming' in Singapore English. *Language and Speech, 43*, 377–401.

Mirka, D. (2009). *Metric manipulations in Haydn and Mozart: Chamber music for strings, 1787–1791.* New York: Oxford University Press.

Patel, A. D. (2003). Rhythm in language and music: Parallels and diferences. *Annals of the New York Academy of Sciences*, *999*, 140–143.

Patel, A. D., & Daniele, J. R. (2003a). An empirical comparison of rhythm in language and music. *Cognition, 87,* B35-B-45.

Patel, A. D., & Daniele, J. R. (2003b). Stress–timed vs. syllable–timed music? A comment on Huron and Ollen (2003). *Music Perception, 21,* 273–276.

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. *Journal of the Acoustical Society of America, 119,* 3034–3037.

Roach, P. (1982). On the distinction between 'stress-timed' and 'syllable-timed' languages. In D. Crystal (Ed.), *Linguistic controversies: Essays in linguistic theory and practice* (pp. 73–79). London, UK: Edward Arnold.

VanHandel, L. (2009). National metrical types in nineteenth century art song. *Empirical Musicology Review, 4,* 134–145.