

Commentary on “An Information-Theoretical Method for Comparing Completions of Contrapunctus XIV from Bach’s *Art of Fugue*”

Stefanie Acevedo[1]

Department of Music, University of Connecticut

ABSTRACT: Information entropy can be a powerful tool for analyzing differences in musical style as suggested in Paz et al. (2022). The application of such a tool, however, requires unambiguous operationalization of the variables being measured, and establishment of clear relations between the variables and the distinguishing features of the style. Without these, interpreting entropy, or any other relative measure of difference, leads to no useful conclusions regarding the similarity between style.

Submitted 2022 June 20; accepted 2022 July 10.

Published 2023 August 10; <https://doi.org/10.18061/emr.v17i1.9158>

KEYWORDS: *Entropy, Style, Voice-leading, Operationalization, Bach*

INTRODUCTION

IN “An Information-Theoretical Method for Comparing Completions of Contrapunctus XIV from Bach’s *Art of Fugue*,” Ivan Paz et al. (2022) use relative entropy distribution measurements to compare the strength of various completions of J.S. Bach’s unfinished fugue, composed by Donald Francis Tovey (1931), Davitt Moroney (1989), Zoltán Göncz (2006), and Kevin Korsyn (2016). The study compares the strength of each completion according to the relative entropy (or Kullback-Leibler divergence) between a completion’s pitch distribution (P) and that of an a priori distribution (Q), calculated from Bach’s unfinished version.[2] The aim of the comparison is therefore to “explore to what extent it is possible to ‘measure’ the similarity of reconstruction to the original [Contrapunctus XIV], and thus determine which [completions] are most ‘Bach-like’ in respect of their voice-leading” (p. 3).

This methodology relies on the assumption that musical style is cognitively encoded as information: style is determined by expectations, which are directly influenced by statistical probabilities of musical components, whether they be pitch distributions, durations, chord transitions, etc.[3] When distributions are compared across musical works, information theory provides quantifiable measurements of difference (here measured as relative entropy) that identify whether something is perceived to be more or less like something else. In the case of Paz et al.’s study, a single measurement aims to show whether a musical piece is more or less “Bach-like.”

The authors note that their analysis is an “exploration”—they are aware of the limitations of this methodology for denoting actual perceptual distance between musical styles. The authors measure the relative entropy between the pitch distributions of each voice of Bach’s unfinished fugue (stratified into the standard four-voice Soprano, Alto, Tenor, Bass parts) and compare them to the pitch distribution of each completion’s respective four voices. Two further adjustments are also made to account for pitch duration (that is, the measurement of pitch distribution controlled for each pitch’s rhythmic duration) and pedals in the bass voice (whereby sustained final pedal notes in the lowest voice are eliminated from the measurements).

Table 1. Relative entropies between pitch and duration probability distributions of completions calculated for separate voices and listed by composer (Replication of results shown in Tables 1, 2, and 3 from Paz et al., 2022).



Table 2. Pitch distributions of Contrapunctus XIV fugue subject 1 bass voice (mm. 1-5).

Pitch & Duration	Count
D3	2
Half Note	2
F3	1
Whole Note	1
G3	2
Quarter Note	1
Whole Note	1
A3	1
Whole Note	1

Merely using pitch distributions for measuring adherence with, or similarity to, a musical style is therefore not appropriate. For instance, one can have two melodic utterances with exactly the same pitch distributions (even if controlled for duration) that are based on completely different syntactic and stylistic norms. Figure 1A shows the first five measures of the Bass voice's first subject from Contrapunctus XIV from *The Art of Fugue*. Figure 1A, by comparison, shows a shuffled version of the same segment of music, with the exact same pitch and duration distributions (shown in Table 2). The melody in Figure 1B line is distinctly unlike the first; The apparent modality (G Dorian?), metrical placement of consonant and dissonant pitches, increased disjunct motion, and rhythmic syncopation might lead some to argue this melody to be non-stylistic for a late Baroque fugue theme. Yet, these two melodies have equal pitch distributions, or a relative entropy measurement of zero, and would be rated as most similar to each other based on the methods employed by Paz et al. (2022).

CAUTION: ENTROPY IS COMPARATIVE

We now turn to the second issue: interpretation of the relative entropy measurement between pitch distributions. The authors claim that their measurement can aid in defining "Bachness" to rate "appropriate completion[s] of the work." But how do relative entropies denote style differences? First, we must have a body of work to compare to (the style that is being emulated), and second, we must interpret the meaning of our measurements within some context.

If measuring whether a completion is truly Bach-like, what should be the referent work? Should one analyze all of Bach's output including his non-fugal works, only his fugal output, or a single fugue movement taken from a larger work? Here, the authors use pitch distributions of individual lines from a single movement (the unfinished fugue) as comparisons. This is done without taking into account whether the piece itself is internally consistent, which creates an unreliable comparison distribution for the purposes of determining "Bach-ness." In his study of previous entropy work on musical style, Joel E. Cohen (1962) critiques early analyses due to issues which are also demonstrated in the analyses by Paz et al. (2022). These studies (and perhaps any musical analysis using entropy measurements) fail to meet basic aesthetic and mathematical assumptions for using information theory methods. The aesthetic assumptions are discussed below. The mathematical assumptions require that the studied corpus a) contains all significant utterances that define the style (stochasticity); b) is large enough to be stylistically homogeneous (ergodicity); c) is homogeneous, independent of time of observation (stationarity); and d) is homogeneous in its patterning (Markov consistency) (pp. 155-157). In the current study, the main comparison distribution fails to meet these assumptions due, mainly, to the size of the sample: a single work cannot be stochastic, homogeneous, or stationary when representing Bach's style.

One of Cohen's (1962) necessary aesthetic assumptions holds that information theory methods are an attempt to quantify listeners' perceptions of stylistic difference (p. 158). This requires understanding the audience's stylistic expectations and that any studied musical sample must a) be representative such that expectation measurements toward it will represent the expectation toward the whole (p. 158); and b) contain patterns that are architectonically independent across levels (i.e., a melody that only effects expectation for melodic components, but not texture, or form) (p. 159). As Cohen argues, given the complexity of music, these assumptions may possibly never be met. Though an attempt must be made if the use of information theory methods is to be meaningful.

Paz et al.'s (2022) final interpretation of results is, consequently, considerably challenging. Due to the abstract nature of entropy measures, scholars have argued for the contextualization and relative comparison for the proper interpretation of entropy measurements (Knopoff & Hutchinson, 1983; Snyder, 1990; Margulis & Beatty, 2008). To provide some context for their results, Paz et al. perform relative entropy analyses of soprano lines from four other fugues, including another from the *Art of Fugue* (Contrapunctus I). The expectation is that these comparisons will provide a relative scale for interpretation based on the authors' (and the listeners') preconceived notions of similarity between composers.

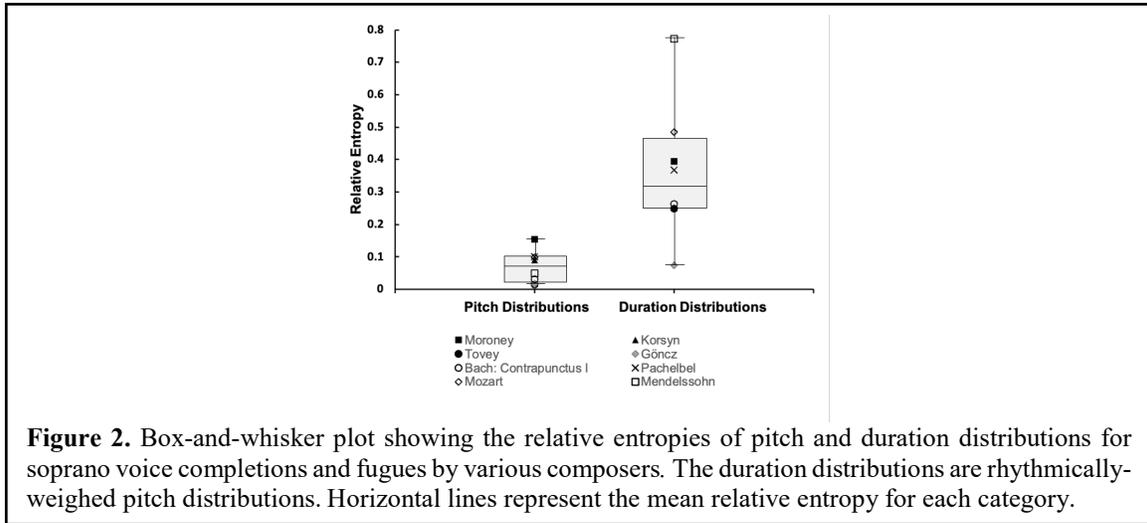
The results (see Table 3) show a wide range of measurements for relative entropy of pitch distributions (in both the completions and other fugues). The scale is much wider between pitch distributions that take into account duration parameters: a range of 0.015-0.101 for pitch distributions versus 0.073-0.773 for duration-controlled distributions. Looking solely at the relative entropy measurements for duration distributions, it is expected that Mendelssohn would differ the most from Bach (the use of chromaticism and various durational variations may account for the difference). In fact, the relative entropies for the durationally-controlled pitch distributions increase as we move forward in time by composer (the increase between Bach and Pachelbel, more-or-less contemporaneous composers, being the exception). However, the duration distribution from Contrapunctus I also shows a fairly high relative entropy (0.262) as compared to the relative entropies of the completions' soprano lines (range of 0.073-0.395). Does this mean that Bach's own output is not stylistically consistent? Does it mean that Göncz, Tovey, and Korsyn write more like Bach than Bach himself?

Table 3. Relative entropies of pitch and duration class probability distributions for soprano voices in completions and fugues by various composers (including Bach's *Art of Fugue* Contrapunctus I). Bolded values show lowest relative entropies. Replication of results shown in Tables 2 and 4 of Paz et al., 2022.

Bach Completions	Soprano Voice Distributions	
	<i>Pitch Distribution</i>	<i>Duration Distribution</i>
Moroney	0.153	0.395
Korsyn	0.090	0.250
Tovey	0.015	0.248
Göncz	0.016	0.073
Other Fugues		
Bach: Contrapunctus I	0.030	0.262
Pachelbel	0.101	0.368
Mozart	0.099	0.486
Mendelssohn	0.049	0.773

Note: The duration distributions are rhythmically-weighted pitch distributions.

The authors go no further in contextualizing the results with regard to relative scale and variability of the entropy measurements. Without this vital step, no conclusions can truly be drawn regarding the similarity between musical examples. One reason for this is the fact that compositional conditions will determine the intrinsic maximum entropy of a style [6]. As Knopoff and Hutchison (1983) argue, assessing entropy (even when employing comparative measures such as relative entropy) is meaningless unless "the precision and comparative validity of entropies [are] determined" (p. 77).[7] They propose using statistical testing to determine whether musical examples are stylistically similar. That is, if the entropy measures (calculated using a specific musical parameter) of two musical examples are significantly different, then it is likely that those examples are not from the same musical style.



The box-and-whisker plot in Figure 2 illustrates the drastic difference between relative entropy measurements in the analysis by Paz et al. (2022). All of the pitch distribution measurements fall within a range of 0.2 units (with 50% of the measurements falling within a range of 0.1). While scale is important to consider, it appears there is little variability in this group of measurements compared to the durationally-controlled pitch distributions, which span a range of 0.8 (with 50% of the measurements falling within a range of 0.25 units) and show much wider variability.[8] It is evident that the variance of relative entropy measures is widely altered by the addition of the rhythmic domain, an important factor to consider when identifying how each musical parameter contributes to identification of musical style. The stark contrast between pitch and duration distributions, and the wide variability of entropy measurements, prevent any quantifiable conclusions to take place.

CLOSING THOUGHTS

Information theory provides a powerful set of tools for analyzing musical expectation. Despite the field's more than 70-year history, the use of entropy as a measure of perceived musical style difference may not lead to significant outcomes until we are better able to operationalize our musical variables. This is vital due to the abstract nature of the measurements. Much like when using decibels to measure loudness, a referent and scale must be determined. Consequently, interpretations must be relative and highly contextualized. As Paz et al. (2022) state, their study was mainly aimed at exploration: a pilot project, of sorts, to employ and test out a tool. Despite its shortcomings, the study provides many points for discussion, and warnings to heed, for those engaging in entropy-guided analysis.

ACKNOWLEDGEMENTS

Thank you to Ed Large for feedback on and Daniel Shanahan for editing this manuscript. This article has been copy edited and layout edited by Jonathan Tang.

NOTES

[1] Correspondence can be addressed to: Dr. Stefanie Acevedo, University of Connecticut, stefanie.acevedo@uconn.edu.

[2] As defined in the original manuscript: “a measure of how one probability distribution differs from another probability distribution which has been taken as a reference” (p. 6). See original for mathematical definition.

[3] See reviews on the history of the use of information theory for analysis in Papadopoulos and Wiggins (1999) and Acevedo and Shanahan (forthcoming).

[4] See Temperley (2007) for a discussion of the statistics of music and analytical applications.

[5] Michael Tenzer (2019) summarizes the contrapuntal (and harmonic) style of the 16th-19th centuries through seven main musical components: melodic motion, relative motion (between voices), (motivic) permutation, consonant and dissonant intervals, dissonance processes (i.e. resolutions of non-harmonic tones), melodic structuring, and interpart relationships (p. 625).

[6] For instance, if comparing pitch entropy (assuming 12-tone equal temperament), a pentatonic melody will have a smaller possible maximum entropy than a heptatonic melody due to the number of available pitches in the alphabet. See Knopoff and Hutchinson (1981) for a discussion of alphabet length effects on maximum entropy calculations.

[7] As a distance measure, the Kullback–Leibler divergence allows us to compare distributions arising from systems with different maximum entropies. However, interpreting the distance measure (relative entropy) still requires contextualization: while a measure of 0 reflects no difference in distributions, there is no upper-bound or referent range for interpretation.

[8] The data provided does not allow for proper significance testing between relative entropies.

REFERENCES

Acevedo, S., & Shanahan, D. (forthcoming). A History of Information Theory Applications to Musical Data. *The Oxford Handbook of Music and Corpus Studies*. New York: Oxford University Press.

Cohen, J.E. (1962). Information Theory and Music. *Behavioral Science*, 7(2), 137-163. <https://doi.org/10.1002/bs.3830070202>

Drabkin, W. (2001). Part-writing. *Grove Music Online*. <https://doi.org/10.1093/gmo/9781561592630.article.20989>

Göncz, Z. (2006). *Contrapunctus 14 für Orgel aus der Kunst der Fuge*. Stuttgart: Carus.

Knopoff, L., & Hutchinson, W. (1981). Information Theory for Musical Continua. *Journal of Music Theory*, 25(1), 17-44. <https://doi.org/10.2307/843465>

Knopoff, L., & Hutchinson, W. (1983). Entropy as a Measure of Style: The Influence of Sample Length. *Journal of Music Theory*, 27(1), 75-97. <https://doi.org/10.2307/843561>

Korsyn, K. (2016). At the Margins of Music Theory, History, and Composition: Completing the Unfinished Fugue in *Die Kunst der Fuge* by J. S. Bach. *Music Theory & Analysis*, 3(2), 115-143. <https://doi.org/10.11116/MTA.3.2.1>

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. <https://doi.org/10.1162/comj.2008.32.4.64>

Moroney, D. (Ed.). (1989). *J.S. Bach: Die Kunst der Fuge*. Munich: G. Henle Verlag.

Papadopoulos, G., & Wiggins, G. (1999). AI Methods for Algorithmic Composition: A Survey, a Critical View and Future Prospects. In *Proceedings of the AISB Symposium on Musical Creativity*, Vol. 124 (p. 110-117). Edinburgh, UK: Society for the Study of Artificial Intelligence and Simulation of Behavior.

Paz, I., Knights, F., Padilla, P., & Tidhar, D. (2022). An Information-Theoretical Method for Comparing Completions of Contrapunctus XIV from Bach's *Art of Fugue*. *Empirical Musicology Review*, 17(1), 2-10. <https://doi.org/10.18061/emr.v17i1.7544>

Snyder, J.L. (1990). "Entropy as a Measure of Musical Style: The Influence of a Prior Assumptions." *Music Theory Spectrum*, 12(1), 121-160. <https://doi.org/10.2307/746148>

Temperley, D. 2007. *Music and Probability*. Cambridge, MA: The MIT Press. <https://doi.org/10.7551/mitpress/4807.001.0001>

Tenzer, M. (2019). Polyphony. In A. Rehding & S. Rings (Eds.), *The Oxford Handbook of Critical Concepts in Music Theory* (pp. 602-647). New York: Oxford University Press. <https://doi.org/10.1093/oxfordhb/9780190454746.013.32>

Tovey, D.F. (Ed.). (1931). *The Art of Fugue (Die Kunst der Fuge) [of] J.S. Bach*. London: Oxford University Press.