

## Tempering the clavier: Examining the intervallic content of Bach’s *Well-Tempered Clavier* through the lens of historical temperaments

Andrew Blake<sup>1†</sup>

<sup>1</sup>Eastman School of Music, University of Rochester, Rochester, NY, USA

† Corresponding author: [ablake10@u.rochester.edu](mailto:ablake10@u.rochester.edu)

Published 16 December 2021; <https://doi.org/10.18061/FDMC.2021.0004>

Author video presentation and/or other conference material: <https://doi.org/10.17605/OSF.IO/UGW86>

### Abstract

While it is understood today that J.S. Bach’s *Well-Tempered Clavier* was intended for a system of “well temperament,” which (if any) specific temperament remains unknown. Excerpts from the work have been analyzed in the context of Werkmeister-III (Gann, 2019), but also within other tuning frameworks (Lehman, 1995). This study addressed the question of temperament from a computational standpoint – in other words, the aim was to determine whether a certain temperament would be a better “fit” for Bach’s WTC when frequency of harmonic intervals and deviation from pure (just) intonation are considered. This study examined accented (simultaneous onset) imperfect consonances’ frequencies as a proportion of all accented harmonic intervals in the WTC fugues, weighting each according to its notated duration in the score. The average expected frequency for each interval was calculated on a basis of key (e.g. major third between tonic and mediant) to control for differences in usage arising from tonal function. Intervallic “emphasis,” then, was used as a metric. This study found no significant correlation between intervallic emphasis for Werkmeister-III, and a small yet significant positive correlation for Bach-Lehman. This indicates that Bach was not avoiding less purely-tuned intervals in these fugues, and emphasizing the “color” of each key.

KEYWORDS: *Bach, Temperament, Well-Tempered Clavier, Werkmeister, Bach-Lehman*

### Introduction

Not all harmonic intervals are handled equally in Western music. In the contrapuntal practices of J.S. Bach, for example, consonant intervals are a much larger part of the harmonic language than dissonant intervals (Huron, 1991). For example, the perfect fifth is used more frequently and freely than the minor second. Furthermore, certain combinations of scale degrees are more frequent than others as a consequence of tonal function, to such an extent that difference in the handling of pitch classes has been used to predict the key of an excerpt or work (Temperley, 1999, Albrecht & Shanahan, 2013). As a result of this, one might expect

the perfect fifth between the tonic and dominant in the major scale would likely be emphasized (used more frequently) over that between the mediant and leading-tone.

When considering tuning systems such as quarter-comma meantone, however, some fifths will be more “in-tune” than others. In this system, 11 of the perfect fifths are tuned slightly flat (696.6c), while one is significantly sharp (737.7c). This is a consequence of the quarter-comma meantone system privileging the major third as a consonance: 8 of the 12 major thirds are justly tuned, but four are similarly sharp (427.4c). As a result of these stark differences between keys, it might be expected that composers would be more likely to write in keys which are not affected by the detuned intervals.

Regarding J.S. Bach’s *Well-Tempered Clavier*, the question of which (if any) tuning system intended is not conclusively answered. It is clear that “well temperament” was intended, but this represents a class of temperaments, rather than one specific solution. In well temperament, all keys are considered to be equally viable for composition. Whereas it was previously thought that this work was intended for equal temperament, it is understood today that this was not necessarily the case (Gann, 2019, 90). Rather than relying solely on historical writings, it is possible that a systematic study of Bach’s *Well-Tempered Clavier* might provide insight into the “well temperament” intended for the work. This study examines Bach’s WTC in the contexts of Werkmeister-III (Gann, 2019, 87-101) and the Bach/Lehman (Lehman, 1995) well temperaments, and additionally considers 12-tone equal temperament and quarter-comma meantone (Gann, 2019, 71-74) as controls.

The expectation of this study is that one or more non-equal well temperaments (Bach-Lehman, Werkmeister-III) will better explain the harmonic interval content in the WTC fugues than other tunings (12-equal temperament, quarter-comma meantone). In short, this study addresses the question of whether Bach is using



certain intervals more or less frequently than expected (based on their tonal function) in a manner consistent with their intonation in each key.

## **Method**

### *Scope of Corpus*

The corpus for this study consists of 48 fugues, from Bach's *Well-Tempered Clavier* (WTC). These fugues were already encoded in the CCARH corpus (Sapp, 1984) by David Huron in 1995, and the sic editions were used in lieu of the ossia versions. While Gann (2019, 97-98) analyzes selected preludes from the WTC, the fugues were chosen in this study for their consistency in texture and number of voices. The preludes of the WTC include vastly differing textures, even between the preludes in C major in books 1 and 2. As this study was concerned with the harmonic interval content of Bach's WTC, certain preludes (e.g. Prelude No. 21 in B-flat Major, WTC Book 2) would create significant outliers in cases containing very few harmonic intervals. More specifically, the data of interest in the WTC fugues is the emphasis and de-emphasis of tonal intervals.

### *Tonal Interval ("Scale-Degree Class Pair")*

For the purposes of this study, a "tonal interval" consists of a pair of chromatic scale degree classes - a pitch class space rotated such that the tonic pitch is always 0, the dominant pitch 7, etc. The notation of these tonal intervals always places the lower pitch first, thus (0, 4) and (4, 0) are different intervals (despite membership in interval class 4). This method is used to account for differences in intervallic usage with respect to key: for example, the perfect fifth (0, 7) would be expected much more frequently than (6, 11). Inversional equivalence is not assumed; (0, 7) represents a perfect fifth whereas (7, 0) is a perfect fourth. Additionally, octave equivalence (e.g. M3 = M10) is assumed for this study. Rather than considering all harmonic intervals, this study only examines those which are "accented" - intervals where two pitches are articulated simultaneously.

### *Measurement of Interval Frequency*

Intervallic frequency was first measured as an interval's frequency of use, weighted using the durations of intervals in which two pitches have simultaneous onsets, as a proportion of all intervals used in a piece. For example, the tonal interval (0, 4) has a frequency of 4.716% in the C Major Fugue from WTC Book 1, whereas (4, 8) only has frequency 0.736%. This discrepancy is consistent with the structure of the fugue - while the tonic key is C major, there are a few brief

tonicizations of A minor using the major third E - G#. The spelling of these intervals was considered, thus the interval E - G# would be registered as a major third whereas E - Ab would be analyzed as a diminished fourth. While the frequencies of intervals can be measured in this way, it could be argued that (0, 4) is not more common in this fugue "because it is more in tune," but rather due to the significance of this tonal interval over (4, 8). For this reason, all tonal intervals were measured using "deviation from expected frequency," and were compared to an average frequency across all fugues of the same mode (major/minor). The major and minor mode frequencies were calculated separately to avoid skewing of data, particularly around scale degree 3 (scale degree classes 3 and 4). Using "deviation from expected frequency" as a measure, (0, 4) in the C Major Fugue is actually 3.297% *below* the average for a major key fugue, while (4, 8) is 0.049% *above* average. Similarly, the minor sixth (4, 0) is 0.054% less frequent than expected, which might be indicative of this fugue spending more time in other tonal regions than others.

### *Rationale for Focus on Imperfect Consonances*

While any interval in the chosen temperaments could be explained as an approximation of just intonation (e.g. the 700c fifth to the 3:2 harmonic ratio of 702c), this study is only concerned with two intervals - minor thirds and major thirds - and their respective inversions. The more pure harmonic ratios of these intervals draw greater attention to their tuning when they are significantly mistuned, as might be most evident with an out-of-tune octave. Thirds are also prioritized over fifths in this study due to the lack of voice-leading constraints on their use. Fifths or octaves, on the other hand, are subject to constraints (e.g. avoidance of parallel or similar fifths). Regarding dissonances, some intervals such as the major second have multiple valid tunings in just intonation. The first three degrees of a justly tuned major scale, for example, consist of two major seconds tuned at 204c (9:8) and 182c (10:9) - a combination which yields a just major third (5:4) of 386c. Additionally, the tuning systems in question all entail varying degrees of preference for the major third or the perfect fifth, with quarter-comma meantone favoring thirds the most. Pythagorean tuning, which is not in the scope of this study, is the system which prioritizes the fifth *slightly* to prioritize the tuning of most (well temperaments) or all (equal temperament) thirds. For these reasons, only the four imperfect consonances are

in the scope of this study, although they are still measured as a proportion of *all* intervals (including perfect fifths, minor seconds, etc.).

*Hypothesis*

The hypothesis is that for some temperament Bach’s use of tonal intervals in the WTC fugues will emphasize those which are more in-tune and de-emphasize those which are farther out of tune. The statistical test used for this study was Pearson’s R correlation, specifically between “deviation from expected frequency of use” and “absolute deviation from just intonation.” There are 12 combinations of experimental conditions, as there are 4 tuning systems (12-equal temperament, quarter-comma meantone, Werkmeister-III, and Bach-Lehman), which can be examined in the “major” and “minor” key fugues, as well as both major and minor together. The null hypothesis was that there is no negative correlation between Bach’s patterns of interval emphasis and some temperament ( $p < 0.05$ ). If such a negative correlation were found, rejection of this null hypothesis would provide evidence consistent with the hypothesis, as well as a possible clue regarding the temperament Bach might have had in mind.

**Results**

Of the four temperaments examined in this study, two were treated as potential candidates for the tuning of Bach’s WTC – Werkmeister-III and Bach-Lehman. 12-equal temperament and quarter-comma meantone were included as control cases, and it was expected that there would be no meaningful correlation with either. Table 1 provides Pearson correlation values for each of the 12 configurations (3 modes, each considered in 4 temperaments). Any interval which does not occur in *any* fugue of the corresponding mode was removed to prevent these from skewing the results (hence the discrepancy in population size for the “Major” and “Minor” cases).

*12-Equal Temperament*

As expected, there was no correlation between intervallic intonation and Bach’s pattern of intervallic emphasis and de-emphasis. No matter which key is used, every interval is tuned equally.

*Quarter-Comma Meantone*

As with 12-equal temperament, there was no significant correlation between interval tuning and patterns of emphasis. Even though the effect size was larger than that of 12-equal temperament, the p-value indicates that these correlations are not significant.

Table 1. Correlation coefficients and p-values for 12 combinations of experimental conditions.

Temperament	Mode	n	r	p
12-ET	Major	1128	1.946e-6	.9999
12-ET	Minor	1080	2.755e-6	.9999
12-ET	Both	2208	2.361e-6	.9999
Meantone	Major	1128	.0102	.7316
Meantone	Minor	1080	-.0047	.8776
Meantone	Both	2208	.0026	.9042
Werkmeister	Major	1128	.0460	.1224
Werkmeister	Minor	1080	.0486	.1035
Werkmeister	Both	2208	.0478	.0247
Bach-Lehman	Major	1128	.0847	.0044
Bach-Lehman	Minor	1080	.0673	.0270
Bach-Lehman	Both	2208	.0757	.0004

*Werkmeister-III*

Under the “major” ( $p = .12$ ) and “minor” ( $p = .10$ ) key conditions, there is not a significant correlation between interval emphasis and intonation. When both major and minor are considered together, however, the result is a slightly positive correlation ( $p = .004$ ). Considering the lack of a significant correlation in the previous tests (and similarity in effect size to the two insignificant correlations), this is likely a result of the larger population size ( $n = 2208$ ).

*Bach-Lehman*

Under all three experimental conditions, the Bach-Lehman tuning indicates a slight *positive* correlation, with all three conditions significant at  $p < .05$ . While significant (if small in terms of effect size), this result is striking in the valence of correlation – a negative correlation would be consistent with the hypothesis, whereas a positive correlation would imply that Bach is, if anything, *emphasizing de-tuned intervals* in the WTC fugues.

**Discussion**

As the results of this study were not sufficient to reject the null hypothesis, the evidence is inconsistent with the initial hypothesis regarding Bach’s patterns of intervallic emphasis and de-emphasis. While there are some potential confounding factors, the implications of these results – that perhaps Bach was *not* targeting a specific tuning system for the WTC – should also be addressed.

### Fugal Constraints

Many of the possible confounding effects on this study stem from the repertoire selected for the corpus – the 48 fugues from Bach’s WTC. The fugues were selected to reduce any confounding effect of textural difference as can be found in the preludes. While the fugues are much more uniform in texture (contrapuntal with generally 3 to 4 voices), the art of writing fugues introduces many constraining factors to the compositional process. In other words, even if Bach might have emphasized or de-emphasized certain intervals according to their tuning, factors such as voice-leading, imitative counterpoint, and playability (one performer with two hands, as opposed to a choral and instrumental ensemble where each part can be more independent) are all factors which might have outweighed any preferred handling of intervallic intonation.

### Tonal Considerations

While the results of the study are not consistent with the initial hypothesis, the data *are* consistent with the earlier assumption that a tonal interval’s function affects its emphasis or de-emphasis when compared to other intervals. Table 1 displays the expected frequency (as a proportion of total intervallic content) for the top 15 imperfect consonances in major key fugues. All tonal intervals except for the lowest column in Table 1 (the minor third between scale degree classes 6 and 9) are diatonic to the major scale and make up a total of 59.51% of all accented harmonic intervals in the major-key fugues. All perfect consonances, intervals with chromatic pitches, and dissonances are encompassed by the remaining 40.49%. Even among these 14 intervals, the tonic-median scale degree class pair has a total frequency of 13.33%, whereas the second most frequent is median-dominant (4, 7) at 8.70%. In other words, over one third of these imperfect consonances are part of the tonic triad, indicating that the major-key fugues emphasize these consonances over others.

While Table 2 includes the top 15 imperfect consonances from the major-key fugues, inclusion of the perfect consonances would introduce only one interval – the perfect fifth (0, 7) at expected frequency 2.40%. This emphasis of almost all diatonic thirds and sixths over even the most common perfect consonance is likely a result of contrapuntal constraints which apply only to perfect intervals. Whereas successions of thirds are permitted in 18<sup>th</sup>-century contrapuntal practice, parallel fifths are not permitted. Leaps to perfect intervals are similarly constrained when compared to imperfect consonances. Furthermore, counterpoint

which is invertible at the octave transforms fifths into fourths – an interval which is considered dissonant when formed against the bass voice.

Table 2. Top 15 imperfect consonances in the major-key fugues of Bach’s WTC.

SD 1	SD 2	Interval	Expected Frequency
0	4	M3	8.01%
7	11	M3	5.62%
4	7	m3	5.56%
9	0	m3	5.32%
4	0	m6	5.21%
2	5	m3	4.83%
5	9	M3	3.81%
11	2	m3	3.80%
7	4	M6	3.14%
5	2	M6	3.11%
0	9	M6	3.03%
2	11	M6	2.92%
11	7	m6	2.92%
9	5	m6	2.23%
6	9	m3	1.35%

### “Well Temperament” as a Class of Tuning Systems

As Bach does not seem to de-emphasize tonal intervals according to intonation in the Well-Tempered Clavier, it may be the case that the work was intended not for a specific well temperament, but rather for any which would be classified as a “well temperament.” Some tuning systems prioritize just intervals at the cost of a handful of significantly out-of-tune ones. For example, Pythagorean tuning prioritizes the just perfect fifth but includes a “wolf” fifth which is 24c flat. Quarter-comma meantone includes 8 just major thirds, but the remaining 4 are 41c sharp (more than 20c sharper than equal temperament). While these tunings can be rotated to start on a different pitch in order to move the out-of-tune intervals, keyboard instruments lack this flexibility.

A “well temperament”, on the other hand, eliminates these starker contrasts. Whereas the major third in meantone is either in tune or 41c sharp, for example, major thirds in Werkmeister-III vary between 4c and 22c sharp, or 6c and 22c in Bach-Lehman. Consequentially, these intervals remain usable across all keys. If Bach’s WTC were intended for something within this “class” of temperaments, the handling of interval tuning would have been a “solved problem”: all keys are equally usable, meaning there would be no

need to treat certain keys significantly different from the others.

#### *Acceptance of Slight Mistuning (Key Color)*

Bach's decision to not emphasize or de-emphasize intervals based on tuning (either through indifference to the tuning of specific keys or through emphasis of greater tuning deviations) might also have served to emphasize the unique profile of each key. Composers of the late Baroque and Classical style periods would eventually take an interest in this phenomenon of "key color." Christian Schubart, an early Classical composer, described the emotional affect which he considered to be expressed through each of the 24 keys (Dubois, 1983). For example, the key of C Major is described as "pure innocence," whereas D-flat Major is a "leering key, degenerating into grief and rapture" (Dubois 1983, 433). Perhaps an apparent indifference toward or emphasis of out-of-tune intervals, then, is meant to showcase the uniqueness of the 24 keys rather than to homogenize them as the original hypothesis would have expected.

#### **Conclusion**

It is obvious from the title that J.S. Bach's *Well-Tempered Clavier* is intended for "well temperament," but this refers to a class of temperaments rather than one specific system which is today referred to as "well temperament." Preludes from the WTC have been analyzed in the context of Werkmeister-III by some, and others have considered other temperaments to be Bach's own system of tuning (Lehman, 1995). While it was not in use at the time for keyboard tuning, equal temperament is a type of "well temperament," as all keys would be considered viable for composition.

In testing for a negative correlation between emphasis or de-emphasis of tonal intervals (pitch class pairs transposed such that the tonic is 0) and deviation from just intonation across Werkmeister-III and Bach-Lehman tuning, the evidence was not sufficient to reject the null hypothesis. The results indicate that Bach either emphasized out-of-tune intervals slightly (Bach-Lehman temperament) or was seemingly indifferent to the tuning of intervals in the fugues. In either of these cases, it is apparent that the "colors" of the 24 keys are emphasized, rather than homogenized. A suitable follow-up to this study might consist of an experiment in which participants listen to the expositions of these fugues in various temperaments, to measure for an effect of tuning on perceived emotional affect of the music itself.

#### **References**

- Albrecht, J., & Shanahan, D. (2013). The Use of Large Corpora to Train a New Type of Key-Finding Algorithm: An Improved Treatment of the Minor Mode. *Music Perception: An Interdisciplinary Journal*, 31(1), 59-67. <https://doi.org/10.1525/mp.2013.31.1.59>
- DuBois, Ted A. "Christian Friedrich Daniel Schubart's "Ideen zu einer Aesthetik der Tonkunst": An annotated translation." PhD diss., University of Southern California, 1983. USC Digital Library.
- Gann, Kyle. *The Arithmetic of Listening: Tuning Theory and History for the Impractical Musician*. Urbana, IL: University of Illinois Press, 2019. <https://doi.org/10.5406/j.ctvpj7j1r>
- Huron, David. "Tonal Consonance versus Tonal Fusion in Polyphonic Sonorities." *Music Perception: An Interdisciplinary Journal* 9, no. 2 (1991): 135-54. <https://doi.org/10.2307/40285526>
- Lehman, Bradley. "Bach's Extraordinary Temperament: Our Rosetta Stone: 1." *Early Music* 33, no. 1 (2005): 3-23. <https://doi.org/10.1093/em/cah037>
- Sapp, C. (1984). Center for Computer Assisted Research in the Humanities. Retrieved from <http://www.ccarh.org/>.
- Temperley, D. (1999). What's Key for Key? The Krumhansl-Schmuckler Key-Finding Algorithm Reconsidered. *Music Perception: An Interdisciplinary Journal*, 17(1), 65-100. <https://doi.org/10.2307/40285812>