On the prevalence of major thirds in Bach's *Well-Tempered Clavier*

Rémi Peyre

April 30, 2026

Abstract

In a famous 1979 paper, Barnes hypothetized that J. S. Bach's using some kind of unequal temperament should have impact the composer's propensity to use such or such major third, depending on how consonant or dissonant that third was. To that end, he had studied the prevalence of thirds among the *WTC* pieces, though the technical means of the time only allowed for analyzing a limited sample of data. In this work I re-instantiate Barnes' analysis, but this time on the complete set of *WTC* pieces, by using more systematic approach, taking advatage of today's availability of numerous MIDI files. Surprisingly, my conclusions are completely different from Barnes', and do not find any statistically significant consistent difference in the choice of major thirds by Bach.

1 Musical and mathematical convention

As this article lies at the intersection of music and mathematics, I will sometimes need use some mathematical convention that may be a bit bizarre from a musician's point of view.

- The musical heights corresponding to the different keys of the keyboards will be called *pitches*. Here the set of pitches is considered to be discrete, with twelve pitches per octave. When using a numeric representation (especially in my computer codes), I will follow the standard MIDI convention: namely C_0 is represented by the value 12, $C\sharp_0$ by 13, D_0 by 14, ..., C_1 by 24, ..., C_4 by 60, and so on. In all this article, adding an integer number i to a pitch means "considering the pitch that is i semitones higher" (or lower of i is negative). So, for instance, $C_4 + 14 = D_5$.
- To refer to a pitch *class*, we will use the word "note". So, there are exactly twelve notes: C, C♯, D, E♭, E, F, F♯, G, G♯, A, B♭ and B. It has to be noted that, in all this article, enharmonic notes are considered to be fully synonymous: so, A♭ is strictly synonymous with G♯ (though, in practice, I shall avoid speaking of "E–A♭ thirds" as much as possible). The numerical encoding and the arithmetic of pitches are extended to notes, except that, for notes, everything has to be considered modulo 12: so, for instance, the note A is encoded by the number 9; and one has $A + 7 = E$, and $A + 12 = A$.
- A (major) *third* will refer to a set of two *notes* (not pitches!) that are set 4 pitches apart. So, there are exactly twelve thirds: C–E, D♭–F, D–F♯, E♭–G, E–G♯, F–A, F♯–A♯, G–B, A♭–C, A–C♯, B♭–D and B–D♯. Note that, in all this article, only

major thirds will be considered; so, for simplicity I will only drop the "major" adjective. Sometimes thirds will only be referred to by their fundamental note: that is, if I speak of the "E third", this actually refers to the E–G♯ third. When needing a numbering convention, the number of a third will correspond to the number of is fundamental note: so, for instance, the E–G♯ third is referred to by number 4.

- A *signature* refers to the set of seven notes that are distinguished at the beginning of a piece, up to enharmonicity. For instance, the signature corresponding to the tonality of C minor is the set of notes $\{\mathrm{F}, \mathrm{C}, \mathrm{G}, \mathrm{D}, \mathrm{A}\flat, \mathrm{E}\flat, \mathrm{B}\flat\}$. When referring to a signature, we will use the name of the fundamental note of the corresonding *major* tonality: so, the signature corresponding to the C minor tonality will be referred to as "the E♭ signature". The numerical encoding of signatures will be done accordingly: so, for this example we would encode the signature in consideration by number 3. Remember that, in the context of this article, we consider enharmonic notes to be fully identical: so, the signatures of D♯ minor and E♭ minor are consider to be the same (and also the same as the signature of F♯ major), encoded by number 6.

2 Prevalence of a given third in a piece

The leading philosophy of my approach is to define some notion of "prevalence of such or such third" in a given piece which can be made almost completely unambiguous, even for a non-musician, while still fitting as best as possible with what is known of psycho-acoustics.

The prevalence of a given third in a *piece* will be merely the average *in time*[*], in that piece, of the "*intensity*" of that third at such of such moment of the piece. The intensity of a third at a certain moment depends itself of the set of pitches being played at that moment.[†] In the sequel of this explanation, let me call such a set of pitches a *chord*. I will explain how the intensity of a given third is computed for a given chord, illustrating that methodology on the $\{\mathrm{C}_4, \mathrm{E}_4, \mathrm{G}_4, \mathrm{B}_4, \mathrm{C}_5\}$ chord.

For each pitch of the chord, one defines two "numeric series" associated to that pitch, each of which is a function from the set of pitches to the set of nonnegative real numbers, in the following way:

- The "1-like harmonics series" of the pitch p associates the value 1 to p itself, the value 1/2 to $p + 12$, the value 1/4 to $p + 24$, and more generally the value $1/2^o$ to $p + o \times 12$, all the pitches not of the form $p + o \times 12$ (for $o \in \mathbb{N}$) getting value 0. So, in the case $p = \mathrm{C}_4$, the pitches $\mathrm{C}_4, \mathrm{C}_5, \mathrm{C}_6, \mathrm{C}_7, \ldots$ are respectively associated with the values $1, 1/2, 1/4, 1/8, \ldots$, all the other pitches being associated with zero.
- The "5-like harmonics series" of the pitch p associates the value 1/5 to $p + 28$ (which is the pitch closest to the fifth harmonic of p), the value 1/10 to $p + 40$, the value 1/20 to $p + 52$, and more generally the value $1/(5 \times 2^o)$ to $p + 28 + o \times 12$, all the other pitches getting value 0. So, in the case $p = \mathrm{C}_4$, one associates $\mathrm{E}_6, \mathrm{E}_7$,

---

[*] Here time if absolute, not relative: if tempo accelerates, the time spanned by a crotchet in the accelerated part of the piece will be considered to be shorter.

[†] The strength of the pitches being played is irrelevant here, as keyboard instrument of Bach's era did not allow for choosing the strength of a pitch: so that the composer never used any intensity indication in the *WTC*.

Pitch	C_4	$D\sharp_4$	E_4	G_4	$G\sharp_4$	B_4	C_5	$D\sharp_5$	E_5	G_5	$G\sharp_5$	B_5
1-like. h. series	1	0	1	1	0	1	3/2	0	1/2	1/2	0	1/2
5-like. h. series	0	0	0	0	0	0	0	0	0	0	0	0
Coinc. series	0	0	0	0	0	0	0	0	0	0	0	0

Pitch	C_6	$D\sharp_6$	E_6	G_6	$G\sharp_6$	B_6	C_7	$D\sharp_7$	E_7	G_7	$G\sharp_7$	B_7
1-like. h. series	3/4	0	1/4	1/4	0	1/4	3/8	0	1/8	1/8	0	1/8
5-like. h. series	0	0	1/5	0	1/5	1/5	0	1/5	3/10	0	1/10	1/10
Coinc. series	0	0	1/5	0	0	1/5	0	0	1/8	0	0	1/10

Pitch	C_8	$D\sharp_8$	E_8	G_8	$G\sharp_8$	B_8	C_9	$D\sharp_9$	E_9	G_9	$G\sharp_9$	B_9
1-like. h. series	3/16	0	1/16	1/16	0	1/16	3/32	0	1/32	1/32	0	1/32
5-like. h. series	0	0	3/20	0	1/20	1/20	0	1/20	3/40	0	1/40	1/40
Coinc. series	0	0	1/16	0	0	1/20	0	0	1/32	0	0	1/40

Table 1: 1-like harmonic series, 5-like harmonic series ans coincidence series for the $\{C_4, E_4, G_4, B_4, C_5\}$ chord.

$E_8, E_9, \ldots$ with the respective values $1/5, 1/10, 1/20, 1/40, \ldots$, all the other pitches being associated with zero.

Once that being done, one gets the 1-like, resp. 5-like, harmonics series *of the chord* by summing the 1-like harmonic series of each pitch composing it. Table 1 shows what it gives for the $\{C_4, E_4, G_4, B_4, C_5\}$ chord. Morally, the value, at pitch p, of the 1-like harmonic series of the chord reprensents how strong the frequency associated to pitch p is, while the value of the 5-like harmonic series gives the intensity at the frequency associated to pitch p, *shifted by the "syntonic-like" comma* resulting from the fact that the $(p-4)$–p third is not just in general.

Now, the *thirds-like coincidence series* of the chord is obtained by taking, at each pitch, the minimum of its 1-like harmonic series and its 5-like harmonic series. See again Table 1 for an illustration. And finally, the *intensity value* (of the chord) for the (n)–$(n+4)$ third is obtained by summing all the values of the thirds-like coincidence series at pitches corresponding to note $n+4$ (that is, at pitches of the form $n+4+o\times 12$, $o \in \mathbb{Z}$). So, for the example of Table 1, the intensity of the C–E third would be obtained by summing the coincidence series for all the E_o pitches, thus yielding $1/5 + 1/8 + 1/16 + \cdots = 9/20$; while the intensity of the G–B third would be obtained by summing the coincidences for all the B_o pitches, yielding $1/5 + 1/10 + 1/20 + \cdots = 2/5$. (And, here, all the other thirds have intensity zero).

3 Basic data

It turned out that MIDI files available over the web are of very variable quality. I even had the impression that many the files that I could find had actually been generated from a keyboard recording (up to some editing), or even from a sound track. In particular, I encountered many files for which some octave notes were played above the notes of the score: but, when applying my script to such files, it yields to a disaster![‡]. I

[‡] Indeed, my computer code assumes that, whenever a pitch is present in the MIDI file, it shall always have the same strength. Therefore, if there are extra octave notes, as weak as they may be in reality, my computer code will not treat them as if they were extra notes with full strength, which will yield an analysis where the computed prevalence of thirds is considerably inflated!

also encountered many files where the most rapid parts of the piece lacked the regularity assumed by the score, which may corrupt the analysis results...

All is all, the golden standard for MIDI files is to have files generated *directly* for an electronic version of the score. The *Lilypond* software, in particular, allows for that. And happily, in 2026, there are so many music passionates around the world, that, thanks to the internet, I managed to get a *Lilypond* file for *every* score of Bach's WTC! In practice, I used the series of scores typed by Knute Snortum (under the CC-BY-SA copyleft licence), which covered the whole *WTC*, and had the advantage of having being generated according to a homogeneous method. May K. Snortum be hugely thanked! ☺

Here I should mention that even the "golden standard" of *Lilypond*-generated MIDI files is not completely unambiguously defined. First, there may be different versions for a given Bach score[§]. Moreover, even for a given score, relative tempi are not perfectly defined (how much faster in *presto* relatively to *allegro*, for instance?), nor are the renderings of certain ornaments (in particular mordents and fermatas). Plus, they may be a handful of errors by the person in charge of typing the electronic version of the file. These issues will be debated in § 7: but let me already tell that it will turn out that changing the source for the *Lilypond*/MIDI files leads to very stable results anyway ☺.

4 Prevalence values

Using a computer script that I wrote, I computed the prevalences for every major third in every of the ninety-six *WTC* pieces (preludes and fugues being treated as different pieces). Note that, for the sake of convenience, I multiplied all the prevalence values by 10^4, thus getting nicer figures. The raw results are shown in Tables 2 to 5. Note that both thirds and signatures were presented sorted according to the circle of fifths. Each cell was coloured in an intensity of red depending to the prevalence of the corresponding third in the corresponding piece.

It is obvious that the (by far) most important factor in explaining the prevalence of a third in a given piece is its position relative to the signature of the piece: this translates into a red diagonal visible on all figures. Also, a clear difference can be noted between major and minor pieces: for a major piece, the most used thirds, by far, are the thirds lying on the tonic (on the diagonal), the fourth (just below the diagonal) and the fifth (just above the diagonal); while for a minor piece, not only these thirds (which, in this case, will lie on resp. the minor third, the minor sixth and the minor seventh, since the tonic is shifted 3 pitches below the signature) are much prevalent, but also the thirds lying on the tonic (9 pitches above the signature) and on the fifths (4 pitches above the signature). Tabel 6 show what one obtains when one averages the prevalences for the four pieces of each tonality.

All of that is by no means surprising, of course! Now the question is: once you remove these effects of tonality, does it remain a systematic difference between the use of such or such third? The simplest idea to answer that question is to just take the average on all pieces: since there is a perfect balance between all tonalities (and even between the fugue and prelude genres), the composer's propensity to favour or avoid

---

[§] K. Snortum himself provided two conflicting versions for two preludes of the second book, namely the preludes of BWV 870 (prelude no. 1 in C major) and BWV 893 (prelude no. 24 in B minor): in both cases, I opted for the score correspoding to Bach's autograph.

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
BWV846a	737	266	50	0	0	15	0	0	25	106	0	178
BWV860a	74	388	125	33	5	9	0	0	0	10	0	0
BWV850a	8	47	252	202	27	16	11	7	0	0	8	60
BWV864a	0	3	266	428	384	87	3	95	9	0	0	16
BWV854a	16	45	63	355	595	295	179	72	4	23	23	12
BWV868a	0	19	0	50	200	753	554	228	0	89	0	18
BWV858a	0	10	0	4	19	77	129	166	32	35	7	14
BWV848a	5	0	0	10	11	20	102	184	114	5	2	0
BWV862a	0	0	0	0	0	0	0	63	426	378	50	0
BWV852a	80	70	18	0	16	31	22	83	449	738	571	105
BWV866a	39	98	0	11	0	0	0	0	0	203	469	358
BWV856a	115	9	14	14	4	4	8	10	0	17	87	222

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
BWV870a	961	239	193	247	67	6	16	31	24	45	322	529
BWV884a	101	296	244	74	4	50	0	0	0	4	0	0
BWV874a	10	211	443	312	79	21	72	7	0	0	15	14
BWV888a	0	85	434	730	252	19	24	73	0	6	36	18
BWV878a	60	24	6	199	528	432	84	9	84	29	5	5
BWV892a	0	5	52	0	59	208	219	60	25	14	21	4
BWV882a	0	0	7	7	3	77	275	210	40	26	56	15
BWV872a	0	0	7	26	41	87	222	410	307	126	39	37
BWV886a	39	2	7	25	24	0	46	242	422	179	75	24
BWV876a	1	12	11	6	0	0	2	8	113	220	239	29
BWV890a	73	40	67	10	2	13	14	32	55	199	345	259
BWV880a	604	228	127	242	134	0	17	60	44	56	378	877

Table 2: Prevalence of fifths in the major preludes of *WTC*'s books I (upper part) and II (lower part). Fifths are presented according to the circle of fifths, as well as pieces: we first list the piece written in the C major tonality, then the piece in G major, and so on, up to the piece in F major.

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
BWV846b	722	440	161	62	106	0	0	22	17	0	72	296
BWV860b	133	406	242	115	14	21	37	0	0	11	0	7
BWV850b	112	671	975	676	67	48	44	0	0	0	0	0
BWV864b	0	48	365	523	332	69	46	60	0	0	16	18
BWV854b	20	7	5	246	531	258	52	0	45	0	0	7
BWV868b	0	0	0	63	254	766	457	128	70	34	0	0
BWV858b	0	0	0	0	32	198	662	395	144	0	44	17
BWV848b	33	5	10	2	6	17	215	430	328	71	44	3
BWV862b	52	0	8	13	9	0	45	319	530	425	71	54
BWV852b	37	73	6	0	0	9	2	25	216	402	225	25
BWV866b	65	41	48	6	0	8	12	2	22	238	484	229
BWV856b	324	45	43	163	0	0	14	46	0	13	276	448

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
BWV870b	581	337	108	26	26	0	0	0	28	0	12	244
BWV884b	98	411	329	25	42	31	25	0	5	8	30	11
BWV874b	61	478	739	597	167	64	42	37	0	0	9	38
BWV888b	60	37	261	497	467	151	42	73	33	8	10	34
BWV878b	7	4	39	512	811	744	93	113	122	41	10	16
BWV892b	10	34	20	58	398	640	314	80	19	39	43	4
BWV882b	5	16	18	32	100	263	272	245	71	24	15	5
BWV872b	6	0	6	0	51	9	327	659	347	38	26	28
BWV886b	41	26	1	76	99	35	118	256	417	370	128	41
BWV876b	17	26	0	0	0	0	0	24	254	685	625	130
BWV890b	93	46	49	0	0	24	13	5	58	304	492	284
BWV880b	311	94	16	39	18	0	0	33	97	101	264	454

Table 3: Same as Table 2, but for major *fugues*.

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
BWV865a	346	133	24	214	31	21	58	56	30	7	97	155
BWV855a	239	369	67	56	132	164	0	0	0	56	0	34
BWV869a	8	324	396	203	41	115	185	94	0	3	25	3
BWV859a	0	0	147	270	125	0	109	65	30	0	0	5
BWV849a	86	25	23	246	588	347	79	80	533	104	0	8
BWV863a	1	76	10	36	292	335	172	46	202	174	66	6
BWV853a	0	9	29	13	97	575	587	221	30	484	275	111
BWV867a	235	0	9	112	38	0	717	926	331	40	275	565
BWV857a	625	119	0	3	43	13	22	235	583	123	3	396
BWV847a	152	55	8	0	0	0	0	0	112	270	17	19
BWV861a	26	182	269	28	0	33	45	0	64	349	358	38
BWV851a	61	4	236	136	31	0	4	9	4	14	153	244

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
BWV889a	87	60	28	31	110	28	13	0	5	6	32	60
BWV879a	103	169	95	4	49	70	67	8	2	20	11	29
BWV893a	10	99	183	78	19	21	38	12	2	6	19	0
BWV883a	19	41	224	410	191	26	156	221	76	4	5	49
BWV873a	42	15	48	203	377	234	59	81	195	69	1	14
BWV887a	0	14	22	61	105	189	143	76	36	106	77	9
BWV877a	2	2	8	0	33	92	205	90	29	34	41	23
BWV891a	46	0	5	17	13	57	340	373	185	31	134	140
BWV881a	104	0	7	6	27	0	36	191	327	143	27	33
BWV871a	44	84	3	0	12	8	8	54	192	442	88	32
BWV885a	24	396	378	116	4	56	72	13	99	491	594	198
BWV875a	58	30	85	79	22	0	0	15	0	5	109	164

Table 4: Same as Table 2, but for *minor* preludes. I sorted the pieces according to the same order as for major fugues as regards *signatures*: so, first the piece written in A minor tonality, then the piece in E minor, then the piece in B minor, and so on.

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
BWV865b	457	332	111	195	151	55	7	27	43	24	74	392
BWV855b	89	90	112	69	181	41	30	10	20	2	20	17
BWV869b	77	224	364	264	87	168	133	102	20	21	50	39
BWV859b	22	30	202	394	207	217	223	240	78	0	34	46
BWV849b	62	36	43	542	510	275	115	250	382	71	17	30
BWV863b	0	9	9	32	269	407	287	33	225	231	111	22
BWV853b	0	18	37	0	53	300	362	207	70	136	226	34
BWV867b	23	0	24	30	82	98	588	611	385	98	331	277
BWV857b	181	86	11	8	35	18	31	238	420	283	126	188
BWV847b	214	166	44	0	19	14	0	0	257	294	110	46
BWV861b	82	157	204	20	7	28	40	14	48	486	412	239
BWV851b	129	27	267	143	56	3	40	79	10	45	194	260

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
BWV889b	155	174	76	243	159	29	0	5	30	11	20	177
BWV879b	133	307	207	74	66	143	44	7	34	54	21	19
BWV893b	28	203	330	315	80	87	140	86	0	10	18	17
BWV883b	21	57	259	317	280	45	96	55	83	15	20	28
BWV873b	41	6	29	252	382	321	133	137	241	52	0	10
BWV887b	0	18	16	44	180	410	188	100	61	154	89	0
BWV877b	0	23	45	90	138	373	404	200	92	216	216	91
BWV891b	55	14	16	50	59	77	214	481	317	96	156	189
BWV881b	158	27	0	23	48	0	16	190	313	163	56	18
BWV871b	174	244	57	0	23	100	20	42	262	578	141	70
BWV885b	102	135	168	16	0	37	32	12	164	392	531	298
BWV875b	147	71	65	194	31	0	13	23	4	20	305	389

Table 5: Same as Table 4, but for minor *fugues*.

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
C major	750	320	128	84	50	5	4	13	23	38	101	312
G major	101	375	235	62	16	28	15	0	1	8	7	4
D major	48	352	603	447	85	37	42	13	0	0	8	28
A major	15	43	331	544	359	82	29	75	11	3	16	22
E major	26	20	28	328	616	432	102	48	64	23	9	10
B major	2	15	18	43	228	592	386	124	29	44	16	7
F♯ major	1	6	6	11	39	154	335	254	72	21	30	13
C♯ major	11	1	6	10	28	34	217	420	274	60	28	17
A♭ major	33	7	4	28	33	9	52	220	449	338	81	30
E♭ major	34	45	9	2	4	10	7	35	258	511	415	72
B♭ major	67	56	41	7	1	11	10	10	34	236	448	282
F major	339	94	50	114	39	1	10	37	35	47	251	500

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
A minor	261	174	60	171	113	34	20	22	27	12	56	196
E minor	141	234	120	51	107	104	35	6	14	33	13	25
B minor	31	213	318	215	57	98	124	73	5	10	28	15
F♯ minor	16	32	208	348	201	72	146	145	67	5	15	32
C♯ minor	57	20	36	311	464	295	96	137	338	74	4	15
G♯ minor	0	29	14	43	212	335	197	64	131	166	86	9
D♯ minor	1	13	30	26	80	335	390	179	55	217	190	65
B♭ minor	90	4	14	52	48	58	465	598	305	66	224	293
F minor	267	58	5	10	38	8	26	213	411	178	53	159
C minor	146	137	28	0	13	31	7	24	206	396	89	42
G minor	59	218	255	45	3	38	47	10	94	430	474	193
D minor	99	33	163	138	35	1	14	31	4	21	190	264

Table 6: Average of prevalences for all four pieces in all major, resp. minor, tonalities.

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
Book I	118	107	110	124	118	123	133	122	131	127	112	109
Book II	97	100	115	132	120	109	97	107	110	117	122	107
Global	108	104	112	128	119	116	115	114	121	122	117	108

Table 7: Average prevalence for each third, when averaged on resp. all the forty-eight pieces of *WTC*'s book I, on all the forty-eight pieces of *WTC*'s book II, and on the totality of all ninety-six pieces in both books.

such or such third *in absolute terms* (in addition of its position relative to the tonality, I mean) should be the only thing remaining after such an averaging—plus, obviously, some random noise, but which should hopefully not be too large given that we have averaged over a rather large number of pieces. The results of that averaging are shown in Table 7.[¶]

The striking conclusion of Table 7 is... that there is actually nothing striking to be seen! The differences are very moderate; and not quite consistent between the two books: it seems to hint at a purely random effect...

However, to conclude on whether the differences are significant of not, it would be metter to have a statistical model accounting for the random effects. This is the object of the next section.

5 Statistical model

5.1 Rationale for using a statistical model

The approach by naive averaging has three drawbacks:

- The first one is that, if one wants to work on an incomplete collection of scores, the summation approach would not work any more: as it is obvious that the prevalence of thirds is highly dependent on the piece's tonality. (Granted, this is not a real issue for this very specific case).
- The second one, and the most important in my opinion, is that using a statistical model will allow one to get *confidence intervals*, so that we know to which extent differences in use between thirds were just random or due tu Bach's actual preferences. This is especially important if we want to use the prevalence of thirds as a hint for Bach's choice of temperament: saying "the prevalence of thirds *suggests* such-or-such tempparement" is fine, but saying "the prevalence of thirds *rules out* such-or-such choice" is better!
- I should also mention that, when taking naive averages, the value for global use of a given third will be mostly impacted by its occurrences in the pieces where, due to the tonality, the expected baseline for the use of this third is already high: and that influence is probably excessive with respect to what intuition suggests. For instance, in the tonality of B major, one expects that the F–A third should virtually never occur. If, nevertheless, the composer chooses to use it quite a couple of time, this should be a rather clear hint that he liked the F–A third in

---

[¶] I used a different coloring scale here, because the coloring scale for Tables 2 to 6 were devised to show differences ranging over several orders of magnitudes: the magenta coloring scale, on the contrary, is directly linear.

an especially strong way! But, in *absolute* terms, such a (relatively) important use of the F–A third in B major piece would be barely noticeable...

These are the reasons why I have found it better to devise a statistical model for analyzing the above results.

5.2 The model used

The statistical model that I am using has 40 parameters, all belonging to $\mathbb{R}_+^*$:

- Twelve parameters $\beta_0^{\mathrm{maj}}, \dots, \beta_{11}^{\mathrm{maj}}$ who say respectively, when a piece is written in the tonality of X-major, what should be the typical prevalence of thirds of X type, resp. of $(X+1)$ type, resp. of $(X+2)$ type, etc. So, if a piece is written in F major, the (approximate) prevalence of C $--$E thirds for such a piece will be decribed by β_7^{maj}.

- Twelve parameters $\beta_0^{\mathrm{min}}, \dots, \beta_{11}^{\mathrm{min}}$ who describe likewise the typical prevalence of thirds of resp. X type, $(X+1)$ type, etc., when the piece is written in the *minor* tonality relative to X-major. (So, the index i in β_i^m is always relative to the key signature, regardless of whether the tonality m is major or minor). For instance, the prevalence of C–Ethirds in a piece of C-minor tonality will be described (as a first approximation) by β_3^{min}.

- Two parameters α_{prel} and α_{fug} which describe the fact that the genre of a piece may make it include, altogether, more often chords including thirds. We set by convention $\alpha_{\mathrm{prel}} = 1$[‖] (so, this is not an *actual* parameter), so that in practice, α_{fug} says how much thirds a fugue tends to have *relative to a prelude.*

- Twelve parameters $\delta_{\mathrm{C}}, \delta_{\mathrm{D}\flat}, \delta_{\mathrm{D}}, \dots, \delta_{\mathrm{B}}$ which describe respectively the propensity of thirds resp. C–E, D♭–F, ..., B–E♭to be over- or underused, in addition to the effects due to the genre and tonality. Here the δ_{Thirds}'s are defined *relatively to the average propensity of using the different thirds*: more precisely, one imposes the geometric average of the δ_{Third}'s to be equal to 1.

- Two important technical parameters γ_{prel} and γ_{fug} called the "granularities" (resp. relative to preludes and fugues). Roughly speaking, the parameter γ_{prel} expresses the fact that, given that you are hearing a third of a certain type at a certain moment of a prelude, you should here also that third at other moments—in particular, for the sole reason that each chord has nonzero duration!

Now, the model states that for every piece, for every type of thirds, the prevalences should be independent. For a piece written in the signature of X (say here, to fix ideas, that the piece is written is a minor tonality[**], and that it is a fugue), we "expect" the prevalence of thirds of type $X+i$ to be approximately equal to the product $\delta_{X+i} \times \alpha_{\mathrm{fug}} \times \beta_i^{\mathrm{min}}$; and more precisely, its law should follow something like a "Poisson distribution of granularity γ_{f}", by which I mean, morally speaking, the law of $\gamma_{\mathrm{fug}} X$, where X is a Poisson distribution of expectation $(\delta_{X+i} \times \mu_{\mathrm{fug}} \times \beta_i^{\mathrm{min}}) / \gamma_{\mathrm{fug}}$.

Of course, Poisson distributions are discrete, while here the prevalences that we are modelling may take continuous values. In fact, what I am using is that I set a *likelihood function* whose shape is directly inspired by the Poisson reasoning, but

---

[‖] The prelude genre seemed to be a more natural reference to me, as it is formally free.

[**] So, given the signature, this would be the tonality of $(X+9)$ minor.

which actually works for any walue of the observation. Namely, if the prevalence of thirds of $(X+i)$ type in the considered piece is x, this will correspond to a ln-likelihood, for the parameter of the statistical model, of

$$\frac{1}{\gamma_{\text{fug}}}\big(x(\ln\lambda-\ln x)-(\lambda-x)\big)-\tfrac{1}{2}\ln\gamma_{\text{fug}}, \tag{1}$$

with $\lambda=\delta_{X+i}\mu_{\text{fug}}\beta_i^{\text{min}}$. Note that this formula remains meaningful even if $x=0$ (but it would lose any meaning for negative x).

Once taken into account the two noramlizing constraints on our 40 parameter (viz., the value of α_{prel} and the average of the δ_{Third}'s), the parameter of our model lives in actual dimension 38: and the model is identifiable[††]

5.3 Estimation of parameters

To estimate the value of the parameters, I used a Bayesian approach, starting from an improper prior on the 40-tuple of the logarithms of the parameters, which prior I take uniform on the 38-dimensional subspace of $\mathbb{R}^{39}$ that they live in.

I then sampled the posterior via a Metropolis-Hastings method, relying on a basic random walk which, at each step, changes exactly one of the parameters (modulo re-projecting on the 38-dimensional space[‡‡]). I have written a *Jupyter* notebook in *Python* language implementing all that, that I shall make available once I find the best way to do so (probably via *GitHub*).

6 Results

Table 8 shows the value of the estimates of all the parameters of the model. It essentially confirms what we had found by naive averaging, with the additional security that we are more confident in the relative weighting of the different pieces in estimating the propensity of the composer to use such or such third. In particular, the ratio between the highest estimate and the lowest estimate for the δ_{Thirds}'s (which are resp. for δ_{A} and δ_{C}) is just 1.25: so, even assuming that this difference is statistically significant (see below), it would not hint to a very impressive effect size in any case... In particular, the 1.25 ratio is by no way comparable with the findings of Barnes, who had some prevalence ratios larger than 3! (And, moreover, the tendency of Bach to favour such or such thirds seems to go in the *opposite* direction of Barnes' findings: see later).

An unexpected finding is the fact that γ_{fug} is estimated to be about *twice* lower than γ_{prel}: a musically sensible way to interpret that fact may be that fugues have a structure that is too "rigid" to allow the composer to deviate much from what the tonality would dictate for the prevalence of the different thirds.

We can compare the estimates obtained via the statistical model with the "naive" averages obtained in the previous section. As Figure 9 shows, both results are very close. So, as regards the estimates themselves, using a statistical model does not bring much w.r.t. the simple averaging method.

[††] More precisely, the model is identifiable provided that the set of data includes at least one piece in each tonality, and at least one piece for each genre. When using the *OpenUTC* catalogue in § 7, there will be only fugues in the set of pieces being analyzed: for that reason, in what case one will drop the parameters α_{prel} and δ_{prel}, and impose $\alpha_{\text{fug}}=1$ (which does not affect the meaning of the δ_{Third} parameters), so that the model remains identifiable.

[‡‡] Of course, if the modified parameter where α_{prel}, the projection would just cancel the walk step: so in practice we never make a step on that parameter.

Param.	α_{prel}	α_{fug}	γ_{prel}	γ_{fug}
Estim.	*1*	1.224	52.2	25.4

Param.	β_0^{maj}	β_7^{maj}	β_2^{maj}	β_9^{maj}	β_4^{maj}	β_{11}^{maj}	β_6^{maj}	β_1^{maj}	β_8^{maj}	β_3^{maj}	β_{10}^{maj}	β_5^{maj}
Estim.	459	308	75.4	33.2	42.5	11.9	1.70	7.98	14.8	15.2	37.4	229

Param.	β_0^{min}	β_7^{min}	β_2^{min}	β_9^{min}	β_4^{min}	β_{11}^{min}	β_6^{min}	β_1^{min}	β_8^{min}	β_3^{min}	β_{10}^{min}	β_5^{min}
Estim.	326	175	59.0	140	162	49.2	4.12	14.3	31.8	15.6	39.3	231

Param.	δ_{C}	δ_{G}	δ_{D}	δ_{A}	δ_{E}	δ_{B}	$\delta_{\text{F}\sharp}$	$\delta_{\text{D}\flat}$	$\delta_{\text{A}\flat}$	$\delta_{\text{E}\flat}$	$\delta_{\text{B}\flat}$	δ_{F}
Estim.	0.900	0.912	0.988	1.128	1.052	1.036	0.993	0.995	1.046	1.048	1.016	0.912

Table 8: Estimations (geometric averages of the posterior distribution) of all the 40 parameters of the model.

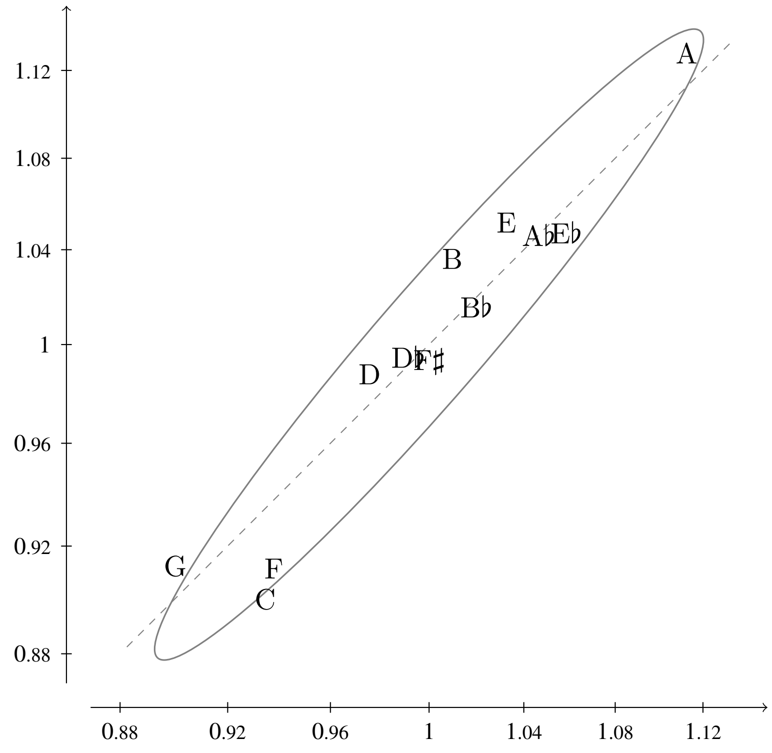

Figure 9: This figure compares the δ_{Third}'s that one would have estimated from a naive averaging (in abcissas) with the Bayesian estimates provided by the statistical model (in ordinates). The ellipse summarizes the first ans second moments of the data: It is the unique ellipse such that the uniform measure inside the ellipse would have the same moments as the data. The dashed line is the "$y = x$" line. The Pearson correlation coefficient r is worth $+96.6\,\%$ $[p < 10^{-6}]$.

However, the main interest of having a statistical model is that we have not only estimates, but also *posterior distributions* for the parameters. Table **??** lists (and illustrates) the main features of the posterior distribution of each δ_{Third}. In particular, the last line looks at whether, for each third T, the possibility that $\delta_T = 1$ (i.e., that Bach was actually just as likely to use that third as the other ones is reasonably "compatible" with the posterior distribution found by the model (in the sense that its quantile, relative to the posterior distribution, is not too close to 0 or 1: in other words, this is a kind of analogue of the p-value in this Bayesian context). Several conclusions may be drawn from these figures:

- Estimation of the δ_{Third}'s is rather precise, as one would logically have expected given the amount of data used: essentially, the 92 % confidence interval for each of these coefficients spans a 1.2 factor.
- Virtually all the Bayesian confidence intervals for the δ_{Thirds}'s are compatible with the possibility that δ_{Thirds} is equal to 1. A notable exception is the $\delta_{\text{A--C}\sharp}$ coefficient, which seems to be significantly larger than 1, with a (bilateral) "p"-value of 1.6 %. Nevertheless, keep in mind that, when proceeding to 12 tests, it rather expected to have at least one "significant" value, even in the case there is *actually* nothing to see: so, this observation concerning the A–C♯third should be taken with much circumspection—all the more that there is a priori no reason why the A–C♯third (which, in baroque temperaments, was neither very just nor very large) should have a special status!
- More striking in the fact that all the three white key thirds (viz., C–E, F–A and G–B) seem (though with little statistical significance) to be slightly underused. Musically speaking, this seems more likely to be voluntary form the composer, as these thirds do have indeed a special status: in most Baroque temperament, these were made just! However, the (very) surprising part is that these thrids appear as *under*used: which is actually the *opposite* direction w.r.t. what one would expect, as well as w.r.t. Barnes' findings! Is it really something deliberate by J.-S. Bach or just a random fluctuation? We will investigate this later, but for the time being, observe that, in Table 6, we found that the case wher·e a third was empirically most used for a given tonality was... the C–E third when writing in the tonality of C major! Which, obviously, does not seem quite compatible with a possible intent of Bach to *avoid* that third...[*]
- But, overall, the general conclusion seems to be that there is little reason to believe that Bach had consistent preferences for or against some specific thirds—at least, when writing the *WTC*.

Investigating the situation third by third is a first natural idea; but, now that we have found that there was only very little to be seen, a natural question is to know whether it is *globally* plausible that Bach had no consistent preferences at all

---

[*] On the other hand, one *might* imagine that Bach somehow "over-compensated" for the time's preference for just thirds. In the cases where the tonality that he was writing in did not explicitly required a heavy use of the C–E, F–A and G–B thirds, *maybe* that Bach was especially careful to let the piece show how the corresponding unjust thirds sounded, and for that reason, *maybe* he was especially careful not to give the impression that he would "relax" toward the just thirds every time that he could... However, this reasoning is quite far-fetched, all the more that it only relies on not-so-significant data: so, do not take the above hypothesis too seriously! ☺

Third	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
$\mathbb{E}_{\mathrm{geom}}(\delta)$	0.900	0.912	0.988	1.128	1.052	1.036	0.993	0.995	1.046	1.048	1.016	0.912
$\mathrm{Var}^{1/2}(\ln \delta)$	0.054	0.054	0.052	0.049	0.052	0.050	0.053	0.053	0.051	0.052	0.053	0.054
$q_{4\%}(\delta)$	0.817	0.828	0.902	1.034	0.960	0.948	0.903	0.906	0.958	0.954	0.927	0.829
$q_{96\%}(\delta)$	0.988	1.003	1.086	1.228	1.150	1.131	1.090	1.091	1.144	1.145	1.113	1.002
"p"	5.3 %	8.8 %	80 %	1.6 %	34 %	49 %	89 %	92 %	38 %	36 %	77 %	8.6 %

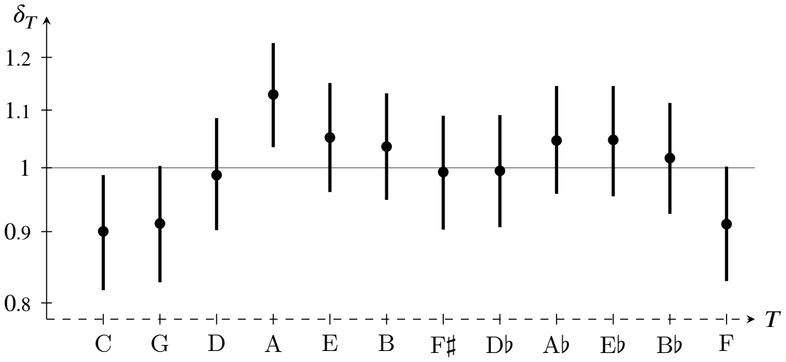

Table 10: The table shows, for each third T, the main characteristics of the posterior distribution of the δ_T coefficient: namely, its geometric expectation (which we used as our estimator), the standard deviation of its logarithm (which in this context, is very close to the standard deviation itself, though mathematically more relevant), its 4 % and 96 % quantiles (which are the bounds of the centred 92 % fluctuation interval), and the "p-value" indicating the position of 1 w.r.t. that posterior distribution. As regards the last line, "p" is defined as the maximum value α such that the centred $(1-\alpha)$ confidence interval of the posterior distribution of δ_T contains the value 1. The figure in the lower part of the table helps visualizing the first, third and fourth lines of the table.

$\times 10^{-5}$	C–E	G–B	D–F♯	A–C♯	E–G♯	B–D♯	F♯–A♯	D♭–F	A♭–C	E♭–G	B♭–D	F–A
C–E	290	−29	−27	−23	−30	−26	−25	−27	−23	−32	−25	−23
G–B	−29	296	−24	−18	−19	−33	−24	−29	−22	−26	−34	−38
D–F♯	−27	−24	271	−22	−22	−23	−25	−23	−31	−23	−23	−27
A–C♯	−23	−18	−22	240	−25	−21	−23	−21	−19	−21	−23	−24
E–G♯	−30	−19	−22	−25	266	−16	−22	−21	−31	−22	−27	−31
B–D♯	−26	−33	−23	−21	−16	254	−26	−21	−21	−22	−20	−26
F♯–A♯	−25	−24	−25	−23	−22	−26	282	−19	−26	−35	−29	−27
D♭–F	−27	−29	−23	−21	−21	−21	−19	277	−28	−22	−33	−34
A♭–C	−23	−22	−31	−19	−31	−21	−26	−28	259	−23	−18	−18
E♭–G	−32	−26	−23	−21	−22	−22	−35	−22	−23	270	−23	−22
B♭–D	−25	−34	−23	−23	−27	−20	−29	−33	−18	−23	276	−21
F–A	−23	−38	−27	−24	−31	−26	−27	−34	−18	−22	−21	291

Table 11: The variance-covariance matrix of the posterior distribution of the 12-dimensional vector $(\ln \delta_T)_T$. Note that, since $\sum_T \ln \delta_T$ is bounded to be zero, the total of the matrix coefficients has to be zero too.

concerning major thirds (in which case the findings above would be nothing more than false positives due to multiple testing), or not. So, we want to make a kind of Bayesian test for the hypothesis that $\vec{\delta}_{\llbracket 0,12 \llbracket} = 1$. Here beware that the δ_{Thirds} *cannot* be independent for the a posteriori distribution, because we know that their geometric average has to always be 1. So, it seems a good idea to use something smarter than a "naive" Bonferroni-like correction.

My idea was to look at the posterior distribution of the vector $(\log \delta_T)_{T \in \llbracket 0,12 \llbracket}$—which, remember, lives in the 11-dimensional subspace of 12-vectors whose coordinates sum to zero. Assuming that one can approximate that posterior distribution with a multinormal distribution in dimension 11 (whose first and second moments we estimate empirically: see Table 11), we find that the value corresponding to perfect third homogeneity has a posterior density such that $2 \ln f_{\text{post}}(\vec{0})$ is 17.5 units below the maximum.[†] In dimension 11, the mass of a multivariate normal distribution having such a low density corresponds to the probability that a $\chi^2(11)$ distribution lies beyond 17.5: that probability is 9.3 %, which is only very slightly significant. So, globally the results are compatible with the complete absence of systametic differences between thirds.

Another idea to investigate the empirical differences in the use of thirds is to look at whether the estimated δ_{Third}'s exhibit a general *structure* when following the circle of fifths (as one would expect if the difference in use were due to the thirds being more or less just), or if it rather oscillates "randomly". In particular, when following the circle of fifths, does one observe a sinewave-like structure in the way the thirds prevalence evolve?[‡] To do so, we look at the intensity of the 1-frequency coefficients in the Fourier series of the $\log \delta_{Third}$'s (when following the circle of fifths, I mean[§]).

[†] By the way, it turns out that the variance-covariance matrix of the posteriori distribution of the $\log \delta_T$'s is remarkably close to being isotropic, the order of magnitude of the observed differences with isotropy being explainable by the sole fact that we used a Monte-Carlo sample to estimate that variance-covariance matrix. If we assume that the posterior law of $(\log \delta_T)_{T \in \llbracket 0,12 \llbracket}$ is *actually* isotropic and that we estimate its density accordingly, this time the $\log_{e^{1/2}}$-density underperformance for the $(1, \ldots, 1)$ value is of 17.2 units, with a corresponding p-value of 10 %.

[‡] That question has the advantage of not being dependent on the uncertainty margins on the δ_{Thirds}'s: you just need use the *estimates*! So in particular this method can also be applied to Barne's figures.

[§] If following the thirds in the pitch order, this would instead correspond to the 5-frequency coefficients in the Fourier series.

When making the computations, we find the the 1-frequency coefficients of the Fourier spectrum accounts for 19 % of the total energy, which corresponds to a p-value of 39 %:[¶] so, one cannot say that there is an observable sinewave structure in the data. As a matter of comparison, using Barnes figures, one would have found, for Method I, 29 % of the frequency spectrum in the 1-frequency components of the Fourier spectrum [$p = 25\,\%$], and for method II, 39 % of the energy spectrum [$p = 11\,\%$]. So my results are less sinewave-shaped than Barnes', even though none of the three series of data has a statistically significant sinewave shape.[‖]

A variant of the above idea would be to look at the real Fourier component *culminating at the third* C–E, since this is near the C–E third that we expect thirds to be the narrowest (resp. near the F♯–A♯ third that we expect thirds to be the widest). This yields somehow neater results (as the near-totality of the energy for the 1-frequency components, amounting to 18 % of the total energy spectrum, belongs to that real component culminating at C–E); however, we are still far from any significance, with $p = 17\,\%$ (as a matter of comparison, for Barnes' data, looking only at that real Fourier component would yield resp. $p = 15\,\%$ for Method I and and $p = 5.3\,\%$ for Method II).

Finally, a last test to investigate the hypothesis that Bach favoured or aboided major thirds in a consistent way consists in comparing the estimates for the δ_{Third}'s when using only the pieces of Book I, resp. only the pieces of Book II. If Bach had really been consistent in his choices, the estimates should be rather close... A visualization of the results obtained when performing this comparison is shown in Figure 12. And this is the final nail in the coffin of the hypothesis of a consistent difference in the use of thirds, as there is absolutely nothing to see: in fact, there are even some blatant inconsistencies:[**] in particular, the F♯–A♯ third, which appeared to be the most favoured in Book I, is the second most *avoided* third in Book II!

7 Robustness of the method

When retrieving MIDI and/or Lilypond files over the internet, it quickly appeared that, from source to source, the contents of MIDI files for a given piece varied in a not completely negligible way[††]. The main differences lies in the scores themselves: most *WTC* pieces are known to exist in different variants (a famous example being the so-called Schwencke measure in BWV846a), either very minor or more impactful: and, from source to source, the variant chosen is not the same. Also, some computer typographs may have added some indications aimed at imporving expressivity of the

---

[¶] Do not forget, in the computation, that the values are constrained to have zero total: so, there is one degree of freedom to remove for the total energy.

[‖] Here one is limited by the fact that nothing says that, asusming an actual difference in the use of thirds, one should have *exactly* a sinewave form: which, given that one only has 12 data value, makes harder to detect something significant—all the more that data are polluted by noise in a nonneglible way!

[**] But, beware that these "inconsistencies", according to the rest of our analysis, are in fact nothing more than random fluctuations: so, assuming that Bach had no consistent preferences, there is actually noting bizarre in observing such inconsistencies! ☺

[††] Here I speak of *actual* changes, not *technical* ones. For instance, globally accelerating the tempo of a given piece would not affect its analysis at all, nor would it change anything to present simultaneous events in a different order, or to consider that such musical phrase should be included in the upper track rather than in the lower track, etc.

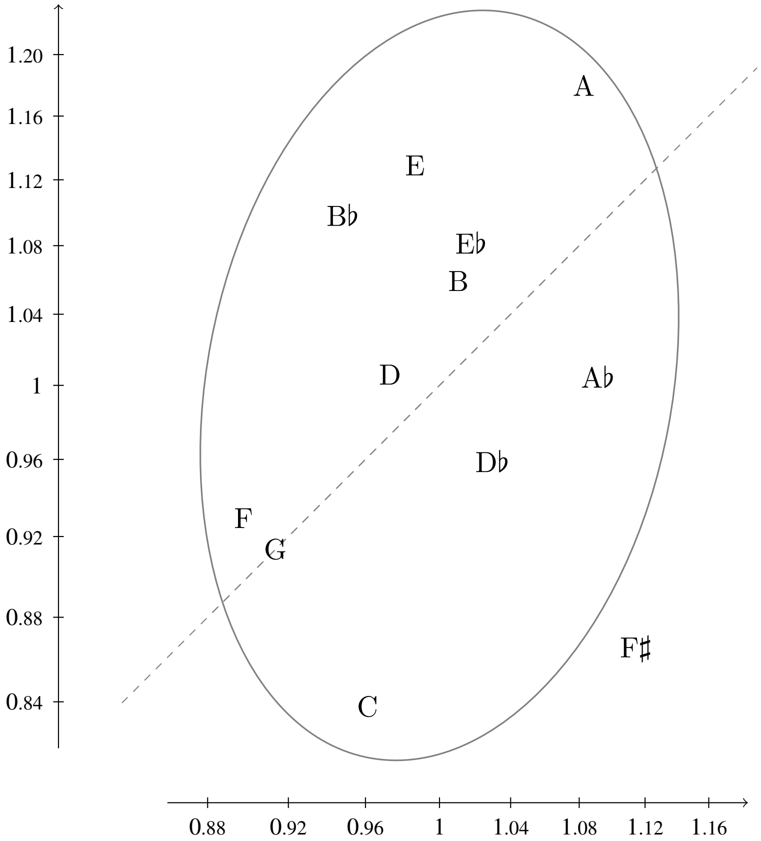

Figure 12: Estimates for the over- (or under-) prevalence of thirds in *WTC*'s Book I (abscissas) vs Book II (ordinates). The Pearson correlation coefficient is +18 % [$p = 28$ %].

piece, like for instance *rallentando* indications, which are certainly relevant from an execution point of view, but were technically absent from Bach's scores[‡‡].

Anyway, *even for an exactly identical score*, the MIDI rendering may differ, because of the ambiguity in the way to interpret some qualitative ornaments like a mordent, a staccato or a fermata, or tempo changes like *allegro* or *andante*. An even more striking feature lies in Knute Snortum's choice of using so-called "articulation" features for his MIDI rendering: this leads to most keys being released just a bit before the end of the notes, leading to a substantial change (of about 10 %) in the prevalence of thirds[*].

Ideally, one should go back to Bach's autograph version and ensure that all the MIDI rendering choices are coherent, and as canonical as possible. But such a work would be *much* beyond my available time (and abilities)...

So, instead of that, I just decided to compare how the estimation of the δ_{Third}'s differs when one changes the sources for the *WTC* pieces. Here however one should take a difficulty into account: the different sources that I found did not, in general, include the full *WTC*. More precisely, I compared Knute Snortum's files with two other sources:

- The *Mutopia* project files (37 files[†]).
- The set of fugues available at the open library of the University of Cape Town (ZAF) [all fugues, that is 48 files].

The fact that some pieces are missing is not actually a problem to compare sources, because my statistical method works the same even if some data is missing. However, one should understand an important point: when considering a subset of pieces of the *WTC*, the estimates of the δ_{Third}'s will differ greatly in comparison with the estimates obtained from the integrality of the two books. Indeed, remember that we have been showing that the differences between the δ_{Third}'s are mostly due to "random" effects, because of which king of musical composition or which theme Bach assigned, mostly arbitrarily, to such or such key. So, one should not (and will not) expect a strong consistence between the estimates obtained from the 36 *Mutopia* pieces and the 96 Snortum's pieces: just because we are analyzing different data, and that signal (if any) is completely overwhelmed by noise. However, one *should* find some strong correaltion by comparing the estimates obtained from the 37 *Mutopia* pieces and from the 37 *corresponding versions* in Snortum's catalogue.

The results are shown in Figure 13. Obviously the stability when changing the source for the MIDI files is very good, which is reassuring concerning the validity of the conclusions that we have drawn from Knute Snortum's files!

[‡‡] It is also common to find *piano* or *forte* indications: but these are discarded by my analysis algorithm in all cases, so this does not matter.

[*] However, this change concerns all the prevalences in an essentially uniform way, so that the impact of estimating the δ_{Third}'s is quite slight in the end.

[†] Namely: BWV 846 to 851, 853, 855 to 857, 860 to 862, 865, and 869 to 871; plus the preludes of BWV 854 and 875, and the fugue of BWV 878.

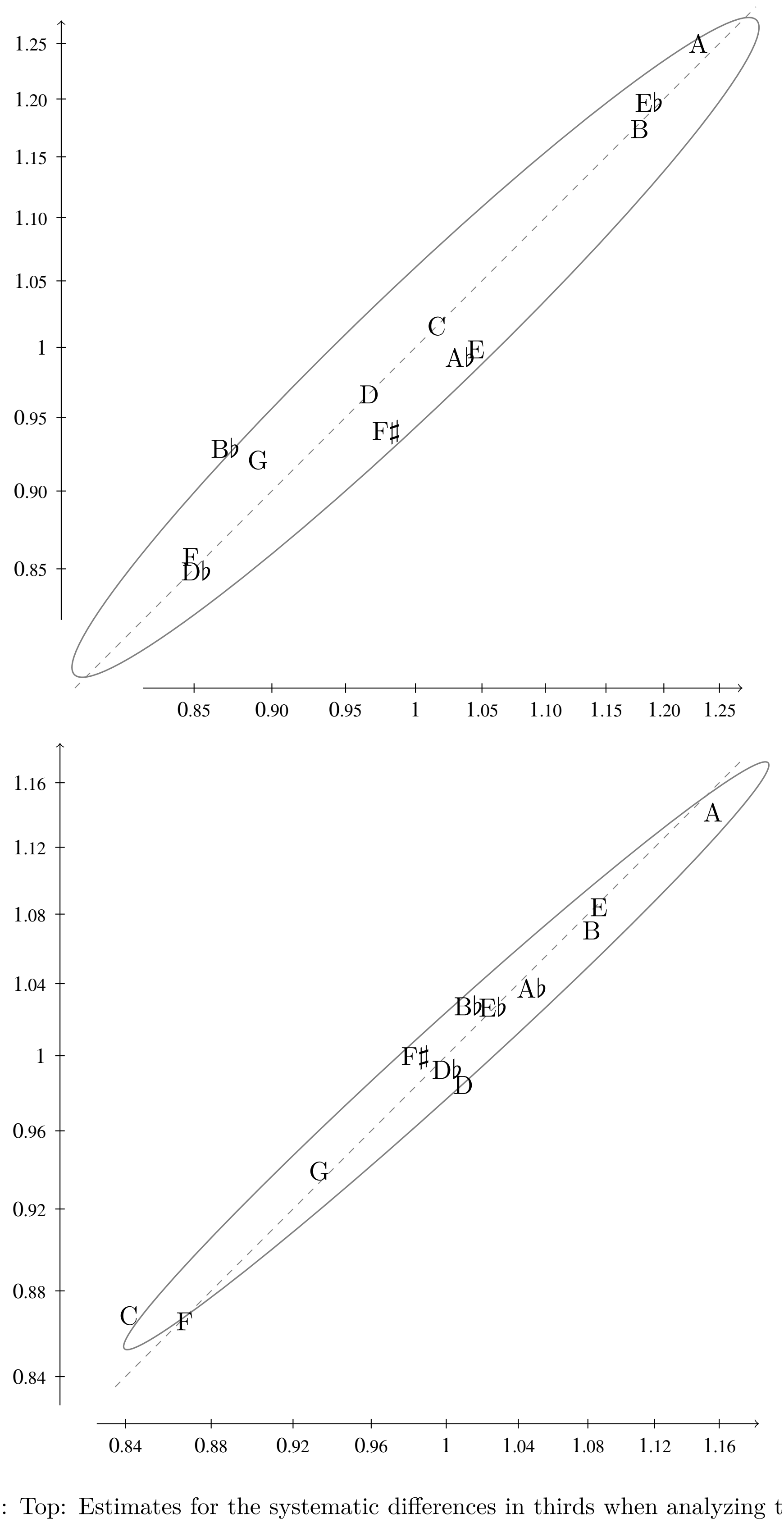

Figure 13: Top: Estimates for the systematic differences in thirds when analyzing the thirty-seven pieces that are common to Snortum's and Mutopia's catalogues, depending on whether we rely on Snortum's versions (abscissas) or Mutopia's versions (ordinates). Pearson's correlation coefficient is +97.0 % [$p < 10^{-7}$]. Bottom: Same when comparing the Snortum catalogue with the OpenUCT catalogue. For that version, the Pearson coefficient is +98.9 % [$p < 10^{-9}$].

8 Future improvements to be done

- Dive deeper in Barnes paper to look at what may explain the differences in our conclusions.
- Add more context and bibliography.
- Explain the rationale behind the modelling choices that I have done and defend them.
- Include the explanations that Pr. Tomita provided me with as regards what is understood by experts on how Bach wrote in pieces and articulated the tonality with the musical content.
- Make the codes (and the raw MIDI files) available on an appropriate deposit.
- Polish the text and the codes.