WHEN VOICES INTERLEAVE: TIMING DEVIATIONS IN SIX PERFORMANCES OF TELEMANN'S FANTASIAS FOR SOLO FLUTE

Patrice Thibaud **Mathieu Giraud** **Yann Teytaut**
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
patrice@algomus.fr

ABSTRACT

Performers convey musical meaning not only through pitch and dynamics but also through micro-timing deviations. This study examines performance analysis and timing in Telemann's 12 Fantasias for Solo Flute, focusing on how musical elements, such as implied polyphony, onset positions, and meter, influence musical performance. We release a corpus with annotations on interleaved voices gathering 11 musicological sources. We first evaluated how simple rules may detect such interleaved voices from the scores. We then analyzed six complete recordings of the fantasias, comparing their timing deviations against a metronomic interpretation. Results reveal significant timing deviations influenced not only by note position within rhythmic groupings but also by the presence of interleaved melodic voices, in particular when these interleaved voices are notated with opposing stems.

1. INTRODUCTION

1.1 Performance Analysis and Timing

Musical performance is a process that extends beyond the mere reproduction of a score. As reviewed by [1], key performance parameters include tempo and timing, dynamics, pitch, and timbre. Additionally, other aspects, such as emotional impact [2], can also be considered. In this study, we focus on *timing*, examining its role in shaping musical interpretation and expressivity. Specifically, *timing deviations* refer to differences in the performed notes' onsets and durations compared to a theoretical, metronomic performance at a constant tempo. Deviations between actual timings and metronomic performances are not random but recurrent [3]. Among existing measures, the Inter-Onset-Interval (IOI) is often used in musical cognition to characterize local variations during a performance at different levels (note, beat, bar, and phrase) [4].

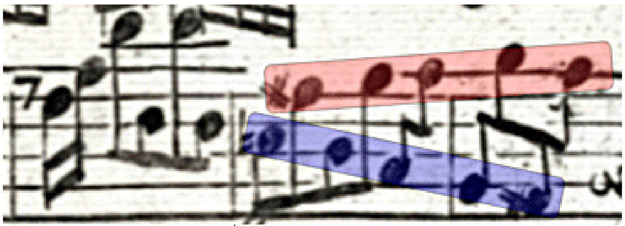
1.2 Timing and Musical Elements

Only a few performance studies have attempted to relate timing deviations to specific structural or compositional aspects of the music. Repp [5] examined piano performances of "Träumerei" by Schumann across different interpretations and measured similarity in the relative durations in consecutive grace notes. He also emphasized extending the initial downbeat as a timing strategy for performing ritards at the ends of melodic gestures in the opening measures of Chopin's "Etude in E Major" [6]. Palmer showed the "melody lead" of corresponding notes in chords played by experimented pianists [7] and studied the internal structure of a trill played in Mozart sonata interpretations [8]. Clarke [9] examined listeners' ability to perceive timing variations in expressive musical sequences, in both tonal and atonal excerpts. The study revealed that a 20 ms lengthening is perceptible within notes ranging from 100 to 400 ms, and that the detectability of a timing deviation varies according to its position within the sequence.

Such timing deviations include the baroque practice of playing *notes inégales*, two consecutive notes of the same rhythmical value with different durations [10]. From its first mention by Loys Bourgeois in 1550 to the late 18th century, such a practice was initially restricted to one-quarter of the *tactus* in duple meter [11]. The strong beats or the first part of each beat were often emphasized compared to the following notes, creating pairs of notes with strong and weak accents. Such an (unequal) performance was generally not mentioned on the scores but theorized in musical treatises and flute methods from this period [12, 13]. Cyr gathered various baroque sources (Quantz, Mattheson) describing where the inequality occurs in baroque dances [14]. Moelants studied the *timing ratio* as defined by Étienne Loulié in 1696 [15], between the longest and the shortest notes when harpsichordists and violinists play *gavotte* [16]. He highlighted the influence of interval size, beat position, and tempo on the *notes inégales*. Honing showed that this timing ratio decreases when tempo increases [17].

1.3 Contributions

How may the performers deviate from the notated durations? Do these deviations follow musical elements, and are they consistent across performers? We aim to investigate how some music elements influence a performance. Focusing on a baroque corpus – the *12 fantasias for solo*



voice annotation	b	u	b	u	B	u	B	u	B	u
note	1	2	3	4	5	6	7	8	9	10

Figure 1. The second episode of the *Vivace fugato* movement of Fantasia No.2 (TWV 40:3) begins at mm.38. It features two interleaved voices, a descending bass line and an upper voice. 🔊 In the autograph score, Telemann occasionally highlights the bass voice in these measures by using stems pointing in opposite directions within the same beam, as here on notes 5, 7, and 9.

flute by Telemann –, we investigate how they may emphasize (or not) some notes, including implied polyphony with interleaved voices. We also investigate whether such implied polyphony can be detected from the scores.

In the following, we introduce the corpus and the annotations of interleaved voices (Section 2) and a method for predicting those annotations (Section 3). We propose a methodology to measure timing deviations in performances and conduct an analysis on six performances, including a focus on interleaved voices (Sections 4 and 5). Finally, we discuss code and data distribution, and the results and perspectives in performance analysis and implied polyphony (Sections 6 and 7).

2. THE TELEMANN FLUTE FANTASIAS CORPUS

The *12 fantasias for solo flute* by Georg Philipp Telemann (*TWV 40:2-13*), edited by himself in Hamburg, probably by 1722-1723, are one of the rare "significant works for unaccompanied flute before the twentieth century" [18]. Each fantasia is a free-form composition, with a sequence of themes evoking improvisation. Telemann's music reflects the mixed taste, "a blending of the French, Italian, English, and Polish national styles" [18]. The fantasias here include different movements which may reflect these influences. Some passages feature implied polyphony through *interleaving* of two or more melodic lines, what Piston describes as a "compound melody" [19]. Such voices [20] may be notated on the Telemann autograph by a beamed group of notes where some stems point downward and others upward (Figure 1).

We built the corpus from MusicXML files transcribed by Árpád Zoltán Szabó [1], and amended these transcriptions by carefully checking the only authenticated copy of the fantasias [21]. Collecting upon 11 musicological sources (Table 1), we gathered annotations about local keys, movements names [2], implied chords and such interleaved voices. We also encoded the harmonic annotations on two fantasias by [20, 22].

Source	mov. names	voices annotations	harmony
Telemann [21]	only tempi	stems indications	–
Brown [23]	partial	–	–
Byrd [24]	–	f7	–
DeBree [25]	–	f7,9-12	–
Eppinger [26]	all	f4-5,7,10-12	–
Hunt [27]	–	f7	–
Kuijken [28]	partial	–	–
Min [22]	all	fugatos only	LK (all), some CP
Porter [29]	all	–	–
Silva [20]	f1-3	f1-3	LK (all), CP (f1-3)
Zohn [18]	all	–	–

Table 1. The musicological sources collected in this dataset analyze some or all of the 12 Telemann fantasias. Harmony annotations include local keys *(LK)* and chords progression *(CP)*.

voices	b u b u	B u B u	u u u u	u b b u	u b u b	B u b u
%	**22.5**	**15.3**	10.8	8.1	**6.3**	**4.5**
voices	b u u b	b m m m	B u u u	b b u u	B u m m	m m m m
%	4.5	3.6	3.6	3.6	2.7	2.7

Table 2. Distribution of main voice annotations in groups of four sixteenths in binary meter movements of the corpus. Bold values show alternating upper/bass voices.

The voice annotations account for 2,324 notes (28.2% of the notes of the corpus) and mostly denote two voices. A middle voice appears in only three movements: the *Allegro fugato* of Fantasia No.7 [27], the *Corrente* of Fantasia No.10, and the *Vivace fugato* of Fantasia No.11 [25, 26]. Voice annotations are encoded by `u` (upper voice, 48% of these annotations), `m` (middle voice, 4.6%), `b` (bass voice, 21.6%), `B` (bass voice, notated with *opposite stem-notes* by Telemann, 25.9%). The sources were largely consistent [3]. Notes without annotations or rare ambiguous cases are tagged by `x`. These elements were labeled using the open-source web platform Dezrann [30] (Figure 2).

Focusing on full annotated groups (*e.g.*, groups with annotation on all notes) of four sixteenth notes within a beat in binary meter, 49.6% of the sixteenth are interleaved, alternating between an upper voice and a bass voice (`b u b u`, `u b u b`, and all combinations with also `B`). In contrast, only 13.5% consist of a single voice, with most being the upper voice and 2.7% the middle voice (Table 2). These groups of alternating notes are mainly (81%) found in the *fugatos* movements (11 of of the 34 movements in the corpus).

[1] `https://imslp.org/wiki/Special:ReverseLookup/236786` (Creative Commons Attribution-ShareAlike 4.0)

[2] Each fantasia is divided into at least three short movements, and up to seven, with the final one always being a dance. The division and naming of these sections were not provided by Telemann but are derived from consensus or majority agreement among the consulted sources.

[3] For example, six opposite stem-notes in the corpus (0.3%) are interpreted as upper or middle voice by some sources. We kept them as bass voices, as in the autograph.

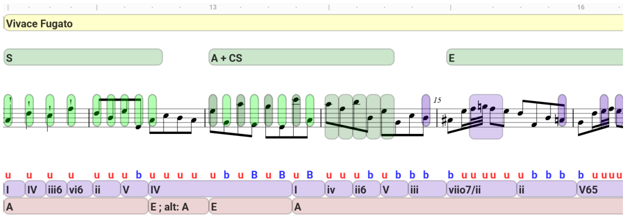


Figure 2. The *Vivace fugato* from Fantasia No.1 (TWV 40:2) mm.11-16 with harmonic labels and identification of the fugato form available on Dezrann 🔊.

3. DETECTION OF INTERLEAVED VOICES IN SCORES

In this section, we study whether, starting from the score, a simple algorithm can detect the interleaved voices as annotated in the corpus. Whereas voice separation in polyphonic music is well studied [31, 32], it is not the case for implied polyphony. Davis [33] proposed a model based on music perception elements to identify where a voice change occurs. Three criteria are applied to analyze all intervals in a piece: (*i*) the size of the diatonic interval [34, 35]; (*ii*) the study of the change of contour direction (ascending and descending voices around the interval) [36, 37]; and (*iii*) the conjunct motion of notes on both sides of the interval [38]. Machine learning models could predict voice changes, but given our small, specific corpus, we prefer to focus on simple rules.

3.1 Methodology

For each interval I between two consecutive notes, several weights are computed:

- w_1 (Pitch distance). Number of diatonic step increments in I, above the third ;
- w_2 (Change of contour direction). Number of changes of direction of the two intervals surrounded I comparing to it (0, 1, or 2);
- w_3 (Stretch of conjunct notes). Sum of the lengths of the stretches of conjunct notes before and after I;
- w_4 (Alternation in intervals). Looking before and after the notes forming I, number of changes of interval direction until this alternation stopped.

Davis proposed the first three weights, w_1, w_2, w_3, and the rule $w_1 > 0$ and $w_1 + w_2 + w_3 \geq 4$ to predict a change of voices [33]. To address the limitations she identified regarding these criteria, we introduce a new weight w_4 and study more combinations of these weights.

3.2 Results

We implemented with music21 [39] these rules and evaluated the results against the existing annotations used as references for all passages with interleaved voices. The original rule achieved a precision of 78% (Table 3), though with a slightly lower recall (64%).

rules	precision	recall	F_1-score
$w_1 > 0$ and $w_1 + w_2 + w_3 \geq 4$ [33]	0.78	0.64	0.70
$w_1 \geq 4$	0.77	0.30	0.43
$w_1 > 0$ and $w_1 + w_2 \geq 4$	**0.81**	0.57	0.67
$w_1 > 0$ and $w_1 + w_3 \geq 4$	0.75	0.39	0.51
$w_1 + w_4 \geq 4$	0.80	**0.72**	**0.76**

Table 3. Classification results on the detection of interleaved voices using several sets of rules.

We achieved better results while considering w_4, the highest recall reaching 72% across the corpus (and 69% when considering only *fugato* movements, data not shown). This model is highly selective. The rule w_1 plays a decisive role. Movements with large intervals between the interleaved voices, such as the Fantasia No. 4 with sixths and tenths, perform better (more than 90% precision). Conversely, the most difficult cases arise when voices are closely spaced, such as in some pedal passages like the opening measures of Fantasia No. 5, where a C pedal is played on off beats.

The remainder of this paper, focusing on timing deviations, relies on the reference voice annotations.

4. PERFORMANCE ANALYSIS METHODOLOGY

4.1 Recordings

There are over 50 recordings of the complete set of the fantasias [40]. We focused here on six performances (Table 4), balancing between two "historical" recordings (Rampal and Kuijken, that were already studied by [40], who evaluated them as two reference interpretations), and more recent ones, as well as balancing between (baroque) traverso and (modern) flutes. As in [41], data were collected for research purposes using spotDL [4] linking Spotify identifiers to YouTube identifiers.

performer	year	instrument	hyperlinks
Jean-Pierre Rampal	1973	flute	YT / MBID
Barthold Kuijken	1978	traverso	YT/ MBID
Amy Porter	2008	flute	YT
François Lazarevitch	2016	traverso	YT / MBID
Emmanuel Pahud	2018	flute	YT / MBID
Katja Pitelina	2020	traverso	YT

Table 4. The six complete audio recordings of the 12 Telemann fantasias analyzed in this study. Recorded during the COVID-19 lockdown, the Pitelana recording is the only non-commercial one and also contains videos. The last column reports, when available, links to YouTube playlists (YT) and MusicBrainz identifiers (MBID).

4.2 Expanded Score Preparation

Some performers, particularly Kuijken, do not strictly adhere to all the repeats noted in the score for 7 of the 12 fan-

[4] `https://github.com/spotDL/spotify-downloader`

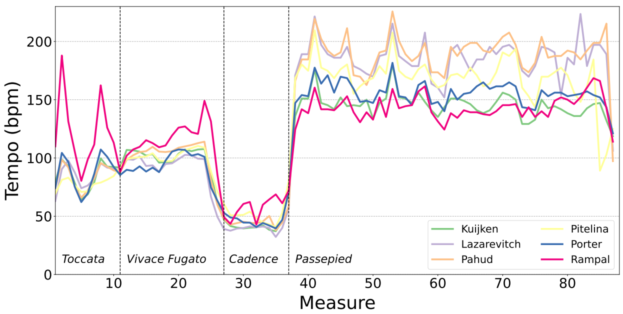


Figure 3. Tempo by measure in *bpm*, computed on three consecutive measures, for the six recordings of Fantasia No.1 (TWV 40:2). All performers underline the four movements of this Fantasia.

tasias. For example, in Fantasia No.2 and No.5, he does not play the final repeat of the last dance movement (a *Gigue* and a *Canary*, respectively). On the other side, Rampal and Pahud, play the Fantasia No.4 with an additional (non-written) repeat in the *Aria da Capo* (the third movement) AABABA, instead of AABA. We listed all these repeats and generated, for each performer, an expanded symbolic score following their performance.

Moreover, performing such baroque music often implies adding ornamentation. Telemann notated only 16 grace notes throughout his entire corpus, most of them in Fantasia No.5, and 64 *tremblements*. Performers may add other ornaments, particularly in slow movements, such as the introductory *Sarabande* of Fantasia No.9, as played by Lazarevitch. Repeats are often ornamented the second time, as abundantly demonstrated by Rampal (for example the *Passepied*, last movement, of Fantasia No.1). We have chosen not to include the written ornaments in the expanded scores, the alignment procedure described below manages to align them with ornamented performances.

4.3 Alignment between Audio and Expanded Score

The symbolic scores are aligned to the audio recordings with a two-fold alignment procedure.

We took into account tuning differences between the traverso and modern flutes with a tuning estimator [42]. Audio spectrograms were computed using 22.05 kHz sampling rate, window size 1024, and hop size 512. Since we deal with monophonic fantasias, we selected the fundamental frequency f_0 as the key audio feature and generated, for each note, discrete harmonic flute templates as Gaussians centered at $\{nf_0\}_{1\leq n\leq 10}$ with $(1/n^p)$-decreasing amplitude [43]. The observation model computes similarities between the audio and the symbolic sequences (note pitches and rests). We further applied a `softmax` function turning these similarities into probabilities, expressed in log-probability space. The few ties (22) were handled by merging the two notes into a single one with summed duration.

Second, we use a *decoding module* relying on Dynamic Time Warping (DTW) [44,45] to retrieve the optimal path (*i.e.*, alignment) between the symbolic sequences from the

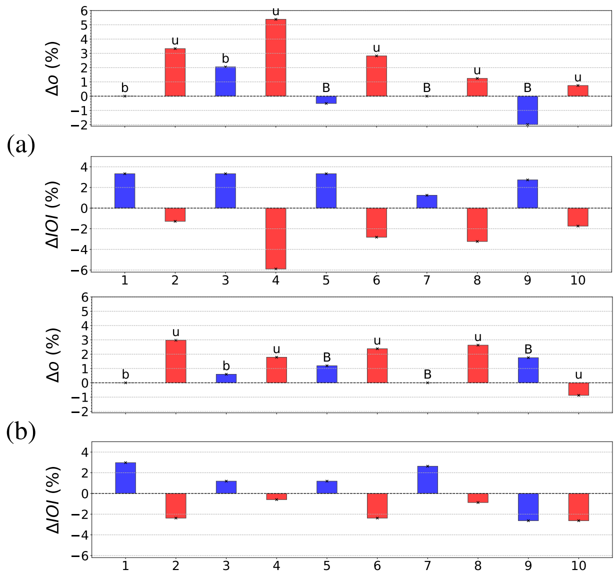


Figure 4. The ΔIOI and Δo values for the passage shown in the Figure 1 are compared across two performances on a modern flute. In Porter's performance (a) 🔊,the average over absolute values of ΔIOI is 2.9%. In contrast, Rampal (b) 🔊, who plays this movement at a faster tempo, has less deviations (1.9%).

expanded score and the output of the observation model. As we do not have ground truth alignment for the fantasias, alignment was not evaluated here but only used as a support for further exploitations, as done in [46] for voice.

Once alignment is complete, each note n has a *performance onset* $o_p(n)$, from which we derive a *performance IOI*, denoted as $IOI_p(n)$. This also allows us to observe tempo variations along each fantasia (Figure 3).

4.4 Metrics for Timing Deviations

We want to compare $o_p(n)$ and $IOI_p(n)$ to theoretical counterparts, assuming no timing deviations. Naturally, performances exhibit global tempo variations (as illustrated in Figure 3), including gradual slowdowns or speedups. However, our focus here is on micro-deviations at a local scale, such as within a group of sixteenth notes. To account for this, we use *each measure* as a reference point for the tempo. For each note n, we thus estimate its *metronomic onset* $o_m(n)$, assuming that the entire measure $M(n)$ encompassing it is played at a constant tempo:

$$o_m(n) = o_m(n_1) + \sigma(n)\big(o_m(n_2) - o_m(n_1)\big)$$

where n_1 and n_2 are the first beats of the measures $M(n)$ and $M(n)+1$, and $\sigma(n) \in [0,1[$ represents the precise *symbolic position* ratio of n within this measure. We also compute the *metronomic IOI* as $IOI_m(n)$. The deviations in IOI and onset timing are then computed:

$$\Delta IOI(n) = \frac{IOI_p(n) - IOI_m(n)}{\mathrm{dur}(M(n))}$$

$$\Delta o(n) = \frac{o_p(n) - o_m(n)}{\mathrm{dur}(M(n))}$$

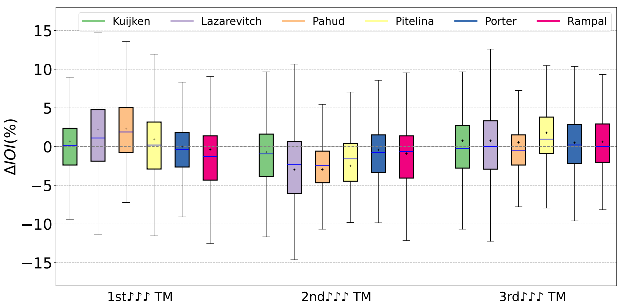


Figure 5. $\Delta IOI(\%)$ of the 1st, 2nd, and 3rd eighth notes within a beat in groups of three, across all movements in a ternary meter and for all performers. The boxes span from the first to the third quartile, while the whiskers extend to the most extreme data points within 1.5 times the interquartile range (IQR). The median is indicated by a horizontal line, and the mean by a '+' sign.

where $\mathrm{dur}(M(n))$ is the duration of the enclosing measure $M(n)$. For example, in the measure 38-39 on the Fantasia No.2 (Figure. 1), Porter plays the second eight note 50ms later than its metronomic position (Figure 4a). This delay corresponds to $\Delta IOI(n_1) = \Delta o(n_2) = 3.3\%$ of the measure duration (about 20% of an eighth note). Note that, by definition, $\Delta o(n_1) = 0$ on the first note (strong beat). The inequality ratio is here 1.30.

5. TIMING DEVIATION RESULTS

We focused on eighth and sixteenth notes. These criteria were analyzed: on-beat/off-beat position, interleaved voice annotation, meter, position of a note within a beat, or a combination of these filters with the notes durations.

Means and standard deviations of ΔIOI are presented in Table 5. These dispersed distributions deviate from normality due to various factors, including human variability in performance, the accentuation of strong beats, and the grouping effects of specific rhythms. This is confirmed using a Shapiro-Wilk test. Because of this non-normality, we use the Mann-Whitney U test [47] to compare two selected populations (Table 6) while simultaneously assessing the effect size. An effect size greater than 0.5 is considered as large. We first focus on generic statistics (Section 5.1), then analyze the impact of interleaved voices (Section 5.2).

5.1 Group of Notes in a Same Beat

In ternary meter, when considering groups of three consecutive eighth notes within a beat (♫♪), the second eighth note is, on average, played shorter by all performers (Fig. 5). Remarkably, Pahud plays the first eighth note of these groups $\Delta IOI_1 = 5.23\%$ longer than the second ones (strong effect size of 0.79), leading to a $\Delta o_2 = 2.35\%$ deviation on the onset of the second notes. According to Garrison, "[Pahud's] approach to articulation is similar to that of traverso players he slurs less often and varies his tonguing style" [40]. For example, in the last movement of Fantasia No.3 (*Gigue* in 6/8 🔊), he emphasizes the first eighth note of each group, particularly on the first beat.

The third notes are played slightly longer compared to the second ones (up to 4.27% for Pitelina), although the effect is less pronounced (0.29-0.43). These deviations align with performance practices in Baroque ternary dances, where "Strong rhythmic articulation, sometimes lead to expressive placement of notes on the second or third subdivision of the beat" [48].

In binary meter, when considering groups of four sixteenth notes within the same beat (♬♬), the first note is, in average, longer. These differences are significant for all performers (except for Rampal). Kuijken plays it 1.78% longer, with a strong effect size of 0.69, as in the first *Vivace fugato* of Fantasia No.3 🔊.

5.2 Interleaved and Opposite-Stem Notes

Interleaved passages most commonly occur in fugato movements, particularly in binary meters within this corpus, such as the two Vivace fugato sections of Fantasia No. 3, separated by four measures of a *Largo*.

We compute here the "♫ interleaved" metrics for all beats in binary meter where a pair of eighth notes appears, each with a different voice annotation (excluding cases with a single voice annotation or an eighth rest). For four performers, there is a significant difference in timing between the first and second eighth notes, with strong effect sizes (0.57-0.67). Remarkably, for three performers (Kuijken, Pahud and Porter), no significant difference is observed when comparing the same situation in non-interleaved passages. Lazarevitch's interpretation is particularly interesting, as he plays the first eighth note 3.14% longer in interleaved passages and 2.64% longer in non-interleaved ones. This could be explained by his consistent approach to inequality in timing within this Baroque corpus, and sometimes a slightly low tempo like in the first movement of Fantasia No.6 (*Adagio*), which he plays in 3'45"🔊, while the fastest performers played it in less than 2'30".

For four performers (Kuijken, Lazarevitch, Pahud, Pitelina), the effects are moderate to strong (0.41-0.60) when one focus on the *opposite stemmed notes* notated in Telemann autograph, as shown on the `B/x` line on Table 6. The specific effect of these notes against the bass notes not annotated by Telemann (`B/b`) is remarkably observed (0.44-0.56) in the same recordings.

The specific effect of interleaved can also be seen on annotated sixteenth-note groups (♬♬), 9.86% of them featuring interleaved voices (Table 2). As in the non-interleaved case, the first sixteenth ΔIOI is quite different from the others, but, notably, there is for Rampal and Pitelina a significant difference (p-value $< 10^{-3}$), between the second and the third sixteenths with strong effect sizes of 0.83 and 0.74, whereas there is no statistical difference in groups without interleaved voices (data not shown). Studies could further investigate these cases, in particular by separating them according to their interleaving patterns, beginning by the upper voice or by the bass voice.

		annotated voices					on beat		off beat	
performer	x	b	B	b + B	u	m	all	♪	all	♪
Rampal	-0.27 ± 9.64	-0.47 ± 6.95	-0.14 ± 5.85	-0.29 ± 6.36	0.26 ± 7.15	-0.60 ± 7.30	-0.53 ± 10.63	-0.28 ± 9.47	0.05 ± 7.46	0.26 ± 9.26
Kuijken	0.07 ± 9.23	-0.17 ± 4.59	0.85 ± 5.46	0.40 ± 5.12	0.52 ± 7.06	-0.68 ± 7.57	0.69 ± 10.40	**1.13** ± 9.19	-0.29 ± 6.75	0.09 ± 8.49
Porter	-0.09 ± 9.78	0.50 ± 5.26	-0.09 ± 5.65	0.17 ± 5.49	0.20 ± 7.06	-0.57 ± 8.22	0.01 ± 10.90	0.88 ± 9.49	-0.07 ± 7.25	0.28 ± 9.00
Lazarevitch	-0.19 ± 10.76	-0.66 ± 5.43	0.26 ± 6.31	-0.14 ± 5.95	**1.28** ± 8.54	**-1.45** ± 8.57	0.79 ± 12.02	**1.55** ± 10.90	-0.73 ± 8.03	-0.90 ± 9.71
Pahud	-0.14 ± 9.15	0.05 ± 4.70	**1.20** ± 5.30	0.69 ± 5.07	0.35 ± 7.42	**-1.26** ± 7.66	0.73 ± 10.43	**1.29** ± 8.81	-0.62 ± 6.59	-0.06 ± 8.20
Pitelina	0.00 ± 9.80	**-1.23** ± 4.87	**-2.32** ± 6.55	**-1.84** ± 5.89	**1.96** ± 8.58	-0.28 ± 7.61	-0.26 ± 11.20	0.45 ± 10.09	0.23 ± 7.38	0.62 ± 9.31

	♫ in binary meter				𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯 in binary meter				♩♩♩ in ternary meter (Fig. 5)		
performer	1st♪ interleaved	2nd♪ interleaved	1st♪ x	2nd♪ x	1st𝅘𝅥𝅯	2nd𝅘𝅥𝅯	3rd𝅘𝅥𝅯	4th𝅘𝅥𝅯	1st♪	2nd♪	3rd♪
Rampal	-0.71 ± 6.28	0.26 ± 7.82	-0.38 ± 10.55	-0.10 ± 10.72	-0.07 ± 3.42	-0.02 ± 3.65	-0.54 ± 4.26	0.24 ± 4.10	-0.36 ± 9.22	-0.91 ± 7.69	0.61 ± 7.03
Kuijken	0.56 ± 4.22	-0.29 ± 4.74	0.76 ± 10.95	-0.42 ± 9.56	**1.23** ± 3.84	-0.55 ± 2.99	-0.60 ± 3.49	-0.15 ± 3.68	0.71 ± 6.10	-0.74 ± 7.40	0.78 ± 9.27
Porter	**1.21** ± 5.35	-0.63 ± 5.35	0.13 ± 10.49	0.43 ± 10.49	0.52 ± 3.31	-0.53 ± 2.94	**-1.14** ± 2.79	0.51 ± 3.41	-0.03 ± 6.52	-0.43 ± 8.02	0.49 ± 7.12
Lazarevitch	**1.72** ± 6.62	**-1.42** ± 6.08	**1.12** ± 11.79	**-1.52** ± 11.00	**1.45** ± 4.05	**-1.01** ± 3.68	-0.16 ± 4.55	-0.65 ± 4.97	**2.16** ± 10.09	**-2.99** ± 8.78	0.75 ± 9.31
Pahud	0.29 ± 4.29	-0.33 ± 4.07	0.01 ± 9.86	0.02 ± 10.24	**1.58** ± 4.19	**-1.06** ± 2.62	-0.90 ± 2.66	-0.34 ± 3.20	**2.28** ± 6.45	**-2.95** ± 4.99	0.55 ± 7.17
Pitelina	**1.21** ± 8.73	-0.75 ± 6.91	0.03 ± 9.91	**1.17** ± 11.12	0.56 ± 3.84	-0.43 ± 3.29	-0.36 ± 4.10	0.67 ± 4.70	0.97 ± 9.12	**-2.50** ± 7.44	**1.77** ± 7.06

Table 5. Mean and standard deviation of the $\Delta IOI(n)$ of notes (excluding rests), in percentage on the length of each measure, for the six performers on all Telemann Fantasias. Values aboves 1% are in bold. (Top). The non-annotated notes (column x) are played with small deviations in contrast to the upper and lower voices (columns u, b +B). (Bottom). Focus on groups of notes within the same beat. Several performers deviate annotated notes written with interleaved voices more than when they are not. They also emphasize the first note of the group.

	Rampal	*Kuijken*	*Porter*	*Lazarevitch*	*Pahud*	*Pitelina*
all the corpus						
b + B/ x	–	0.55	0.54	0.54	0.58	0.44
B/ x	–	0.57	–	0.56	0.60	0.41
u/ x	0.54	0.53•	–	0.57	0.52•	0.58
B/ b	–	0.56•	–	0.55•	0.55	0.44
u/ b + B	0.53•	–	0.46	0.53•	0.43	0.65
all the corpus						
all: on beat / off beat	–	0.57	0.54	0.60	0.61	–
♪: on beat / off beat	–	0.55	0.54	0.61	0.58	–
binary meter						
♫ interl.: 1st / 2nd	–	0.58	0.64	0.67	0.57•	–
♫ x: 1st / 2nd	–	–	–	0.60	–	–
𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯: 1st / 2nd	–	0.69	0.64	0.72	0.75	0.61
𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯: 2nd / 3rd	0.59	–	0.59	0.44•	–	–
𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯𝅘𝅥𝅯: 3rd / 4th	0.40	0.45•	0.28	–	0.43	0.39
ternary meter						
♩♩♩: 1st / 2nd	–	0.58•	–	0.69	0.79	0.63
♩♩♩: 2nd / 3rd	0.42•	–	0.43•	0.36	0.31	0.29

Table 6. Effect size for Mann-Whitney U test measuring the magnitude of the difference between two groups on the metric $\Delta IOI(n)$, focusing on cases where the p-value is $< 10^{-2}$ (•) or $< 10^{-3}$ (all others).

6. DATA AND CODE AVAILABILITY

We provide, under open licenses, curated scores, annotations on structure (movement names, repeats), and annotations on implied polyphony and harmony, under the *recherche.data.gouv.fr* long-term archive at `https://doi.org/10.57745/MSLVWS`. We further distribute these materials, together with code enabling the reproduction of results and figures, via the git repository at `https://gitlab.com/algomus.fr/telemann-voices-interleave`. The dataset also includes note-level synchronization with each recording, along with computed metronomic onsets and timing deviations. These data can further be browsed, and queried on a web application built with Streamlit [5] .

[5] `https://telemannvoicesinterleave.streamlit.app`

7. CONCLUSIONS AND PERSPECTIVES

Whether they play modern flute or traverso, musicians significantly play with timing deviations. Analyzing six performances of the Telemann fantasias, we showed that the initial notes of groups of three eighths in a ternary meter or four sixteenths in a binary meter are emphasized, being longer (up to 5.23% relative to the measure duration) than the other notes in the group.

Interleaved voices are a significant part of the Telemann fantasias – 28.2% of the notes from our musicological sources. We showed that simple rules allow to detect these voices from the score with a F_1-score of 76%. *Notes inégales* are a standard Baroque practice, but we showed here that they are quite significant within interleaved voices. Three of the six performers play with significant timing deviations in such eighth notes, whereas they do not in non-interleaved passages. Notes written with opposite stems in Telemann's autograph score receive particular emphasis.

Any study on performance depends on alignment quality. Many audio-to-score aligners exist today, relying either on audio features (*e.g.*, chromas) or more advanced observation models (*e.g.*, deep learning-based approaches). A natural extension of our work would be to systematically evaluate such methods on annotated ground truth and assess their performance compared to our simple, $f0$-based model. Explicitly considering ornamentations may also better capture the performers' interpretations. The variability of repeats, handled manually here, presents another research opportunity.

The current study was limited on six recordings and may present biais specific to these artists (and this corpus). Given the large number of existing recordings of these fantasias, further studies could consider additional audio data. Finally, the performances could be analyzed on other music elements, such as pitches, intervals, or patterns, or other passages with implied polyphony, such as arpeggios.

8. ACKNOWLEDGMENTS

The authors sincerely thank Florence Levé and Ken Déguernel from the Algomus team for their insightful comments, as well as Emmanuel Leguy for his support with the Dezrann platform. We are also grateful to the conference reviewers for their constructive feedback.

9. ETHICS STATEMENT

The dataset does not contain any listener-specific or personally identifiable information, as it is based solely on previous academic studies.

10. REFERENCES

[1] A. Lerch, C. Arthur, A. Pati, and S. Gururani, "Music Performance Analysis: A Survey," in *ISMIR 2019*, Delft, Netherlands, 2019.

[2] S. B. Dias and L. J. Hadjileontiadis, "Telemann's Fantasia No 2 à Travers Sans Basse: Compositional Intentions, Performance Characteristics and Audio Footprint Signal Processing Analysis," in *Perspectives on Music, Sound and Musicology: Research, Education and Practice*, ser. Current Research in Systematic Musicology, L. Correia Castilho, R. Dias, and J. F. Pinho, Eds. Cham: Springer International Publishing, 2021, pp. 31–47.

[3] Gabrielsson, Alf, "Studying emotional expression in music performance," *Bulletin of the Council for Research in Music Education*, pp. 47–53, 1999.

[4] A. Friberg and G. Battel, "Structural Communication," in *The Science & Psychology of Music Performance: Creative Strategies for Teaching and Learning*, Apr. 2002.

[5] B. H. Repp, "Diversity and commonality in music performance: An analysis of timing microstructure in Schumann's "Träumerei."," *Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2546–2568, 1992.

[6] ——, "A microcosm of musical expression. I. Quantitative analysis of pianists' timing in the initial measures of Chopin's Etude in E major," *The Journal of the Acoustical Society of America*, vol. 104, no. 2 Pt 1, pp. 1085–1100, Aug. 1998.

[7] C. Palmer, "On the Assignment of Structure in Music Performance," *Music Perception: An Interdisciplinary Journal*, vol. 14, no. 1, pp. 23–56, 1996.

[8] ——, "Anatomy of a Performance: Sources of Musical Expression," *Music Perception: An Interdisciplinary Journal*, vol. 13, no. 3, pp. 433–453, 1996.

[9] E. F. Clarke, "The perception of expressive timing in music," *Psychological Research*, vol. 51, no. 1, pp. 2–9, Jun. 1989.

[10] D. Fuller, *Notes Inégales*, ser. The New Grove Dictionary of Music and Musicians. Oxford University Press, 2001, vol. 1.

[11] G. Houle, "V- Articulation of Quantative Meter," in *Meter in Music, 1600-1800: Performance, Perception, and Notation*, ser. Music–Scholarship and Performance. Bloomington: Indiana University Press, 1987.

[12] J. D. Boland, *Method for the One-Keyed Flute, Baroque and Classical*. Berkeley: University of California Press, 1998.

[13] C. A. Fontijn, "Quantz's unegal: Implications for the performance of 18th-century music," *Early Music*, vol. XXIII, no. 1, pp. 54–64, Feb. 1995.

[14] M. Cyr, *Performing Baroque Music*. London: Routledge, 2016.

[15] E. Loulié, *Elements Ou Principes de Musique, Mis Dans Un Nouvel Ordre*. Paris, France: Christophe Ballard, 1696.

[16] D. Moelants, "The Performance of Notes Inégales: The Influence of Tempo, Musical Structure, and Individual Performance Style on Expressive Timing," *Music Perception: An Interdisciplinary Journal*, vol. 28, no. 5, pp. 449–460, 2011.

[17] H. Honing, "Evidence for tempo-specific timing in music using a web-based experimental setup," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 32, no. 3, pp. 780–786, 2006.

[18] S. D. Zohn, *Music for a Mixed Taste: Style, Genre and Meaning in Telemann's Instrumental Works*. Oxford: Oxford University press, 2008.

[19] W. Piston, *Counterpoint*. New York London: W. W. Norton & Company, 1947.

[20] P. da Silva and A. Carlos, "A performance guide to three of Telemann's 12 fantasias for flute," Ph.D. dissertation, University of Alabama Libraries, 2012.

[21] G.P. Telemann, *12 Fantasias for Flute without Bass, TWV 40:2-13*, ser. Royal Conservatoire of Brussels, Library (B-Bc). Hamburg: Telemann, 1732-1733, no. T 5823 W.

[22] Kyoung Joo Min, "Analyses of the twelve fantasies for solo flute(1732-1733) and the twelve fantasies for solo violin(1735) of G. Ph. Telemann :," Ph.D. dissertation, University of Washington, 1998.

[23] R. Brown, "Telemann Fantasias: A feat of ingenuity and inspiration," 2015.

[24] D. Byrd and T. Crawford, "Problems of music information retrieval in the real world," *Information Processing & Management*, vol. 38, no. 2, pp. 249–272, Mar. 2002.

[25] R. de Bree, "Telemann's Fantasias: A Fantastic Source of Inspiration. How can I improve my improvisatory skills , using Telemann's Fantasias ?" 2020.

[26] Eppinger, S., "Georg Philipp Telemann : 12 Fantasien für Flöte solo, Teil I," *TIBIA*, vol. 2, pp. 86–99, 1984.

[27] L. Hunt, "A Horn Transcription of the twelve fantasias for flute without bass, TWV 40:5-8, by Georg Philipp Telemann," Ph.D. dissertation, University of Georgia, 2016.

[28] Barthol Kuijken, "12 Fantasias TWV 40:2-13 (with Facsimile)," 1987.

[29] A. Porter, "Telemann: 12 Fantasias for Flute without Bass – A Study Guide DVD," 2008.

[30] C. Ballester, B. Bacot, L. Bigo, V. N. Borsan, L. Couturier, K. Déguernel, Q. Dinel, L. Feisthauer, K. Frieler, M. Gotham, R. Groult, J. Hentschel, A. d'Hooge, D.-V.-T. Le, F. Levé, F. Maccarini, I. Maričić, G. Micchi, M. Müller, A. Stamatiadis, T. Taffin, P. Thibaud, C. Weiß, R. Yang, E. Leguy, and M. Giraud, "Interacting with Annotated and Synchronized Music Corpora on the Dezrann Web Platform," *Transactions of the International Society for Music Information Retrieval*, vol. 8, no. 1, May 2025.

[31] E. Chew and X. Wu, "Separating Voices in Polyphonic Music: A Contig Mapping Approach," in *Computer Music Modeling and Retrieval: Second International Symposium, CMMR 2004*, Esblerg, Denmark, May 2004, pp. 1–20.

[32] E. Cambouropoulos, "Voice And Stream: Perceptual And Computational Modeling Of Voice Separation," *Music Perception - MUSIC PERCEPT*, vol. 26, pp. 75–94, Sep. 2008.

[33] S. Davis, "Implied Polyphony in the Solo String Works of J. S. Bach: A Case for the Perceptual Relevance of Structural Expression," *Music Perception*, vol. 23, no. 5, pp. 423–446, Jul. 2006.

[34] G. A. Miller and G. A. Heise, "The Trill Threshold," *The Journal of the Acoustical Society of America*, vol. 22, no. 5, pp. 637–638, Sep. 1950.

[35] E. Narmour, *The Analysis and Cognition of Basic Melodic Structures: The Implication-Realization Model*. Chicago: University of Chicago Press, 1990.

[36] G. A. Heise and G. A. Miller, "An Experimental Study of Auditory Patterns," *The American Journal of Psychology*, vol. 64, no. 1, pp. 68–77, 1951.

[37] D. Huron and M. Royal, "What Is Melodic Accent? Converging Evidence from Musical Practice," *Music Perception: An Interdisciplinary Journal*, vol. 13, no. 4, pp. 489–516, 1996.

[38] W. J. Dowling, "The perception of interleaved melodies," *Cognitive Psychology*, vol. 5, no. 3, pp. 322–337, Nov. 1973.

[39] M. S. Cuthbert and C. Ariza, "Music21: A toolkit for computer-aided musicology and symbolic music data." in *ISMIR*, J. S. Downie and R. C. Veltkamp, Eds. International Society for Music Information Retrieval, 2010, pp. 637–642.

[40] L. Garrison, "Telemann's Twelve Fantasias: A Survey of Editions and Recordings," *Flutist Quarterly*, vol. 45, no. 5, pp. 24–31, 2020.

[41] D. Edwards, S. Dixon, and E. Benetos, "PiJAMA: Piano Jazz with Automatic MIDI Annotations," *Transactions of the International Society for Music Information Retrieval*, vol. 6, no. 1, Sep. 2023.

[42] B. McFee, C. Raffel, D. Liang, D. Ellis, M. McVicar, E. Battenberg, and O. Nieto, "Librosa: Audio and Music Signal Analysis in Python," in *Python in Science Conference*, Austin, Texas, 2015, pp. 18–24.

[43] A. Cont, "A Coupled Duration-Focused Architecture for Real-Time Music-to-Score Alignment," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 6, pp. 974–987, Jun. 2010.

[44] S. Dixon and G. Widmer, "MATCH: A Music Alignment Tool Chest," in *Proceedings of the 6th International Conference on Music Information Retrieval*, London, United Kingdom, 2005.

[45] M. Müller, "Dynamic Time Warping," in *Information Retrieval for Music and Motion*. Springer, Berlin, Heidelberg, 2007, pp. 69–84.

[46] L. Salais, P. Arias, C. L. Moine, V. Rosi, Y. Teytaut, N. Obin, and A. Roebel, "Production Strategies of Vocal Attitudes," in *Interspeech 2022*. Incheon, Korea: ISCA, Sep. 2022, p. 4985.

[47] H. B. Mann and D. R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[48] M. Little and N. Jenne, *Dance and the Music of J. S. Bach, Expanded Edition*, ser. Music: Scholarship and Performance. Bloomington, IN: Indiana University Press, 2001.