Annotating Texture and Imitation Patterns in a Corpus of Slow Movements in Corelli's Trio Sonatas

JÉRÉMIE ROUX
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France
Univ. Montpellier, CNRS, UMR 5506 LIRMM, Montpellier, France

MATHIEU GIRAUD
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

FLORENCE LEVÉ
Université de Picardie Jules Verne, MIS, UR 4290, F-80000 Amiens, France
Univ. Lille, CNRS, Centrale Lille, UMR 9189 CRIStAL, F-59000 Lille, France

ABSTRACT: The Corelli trio sonatas are emblematic of the mid-Baroque period, characterized by their elegant melodies, intricate counterpoint, and systematic development of thematic material. Extending harmonic annotations from Hentschel et al. (2021), we provide a dataset with annotations of all 38 slow movements of the *Sonate da chiesa* Op. 1 (1681) and Op. 3 (1689), focusing on this Baroque writing style. The 7,000+ annotations describe the texture (homorhythmy, imitations, suspensions), further detailing thematic and imitation patterns, with other information such as rhythmic density. We provide audio synchronizations to a widely available recording and release the corpus for download as well as through the Dezrann platform.

Corpus : https://doi.org/10.57745/QUZ8KO | dezrann.net/explore/corelli-trio-sonatas

Submitted 2024 July 19; revised November 14; 2025 April 30; accepted 2025 May 7.
In press. https://doi.org/10.18061/emr.vxxix-x.xxxx 10.18061/emr.v20i1.10018

KEYWORDS: *corpus, Corelli, Baroque music, violin, patterns, texture, imitations, harmony*

ARCANGELO Corelli's oeuvre is almost exclusively made up of instrumental works for string ensembles, including Trio Sonatas (*Sonata a Tre*) and *Concerti grossi*, the latter consisting of a dialogue between small groups of soloists and the orchestra. The Italian composer (1653-1713) had a notable influence on his contemporaries and successors (Rothstein, 2006; Sanna, 2013).

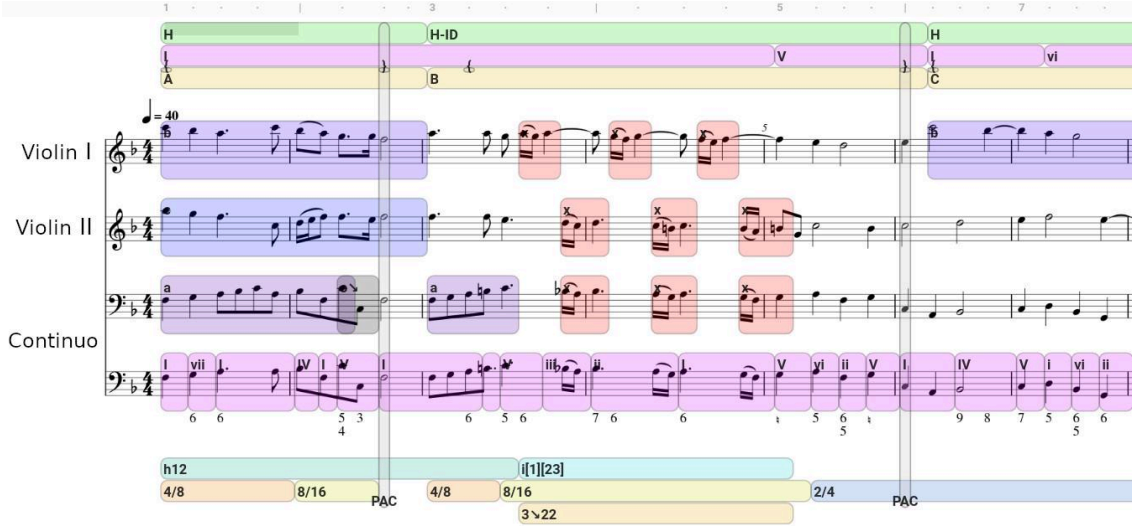
As an outstanding violinist and a renowned composer, Corelli was protected by powerful people (Pincherle, 1954). In Rome, Corelli published five sets of 12 Trio Sonatas, for two violins and a basso continuo, each sonata having about 4 movements: the Church Sonatas Op. 1 (1681) and Op. 3 (1689), the Chamber Sonatas Op. 2 (1685) and Op. 4 (1694), and a mixed Op. 5, including the variation on *La Folia* (1700). An additional collection was published posthumously (WoO 5-10, 1714). In the Trio Sonatas, the violins are equal in their range and their melodic role. Crossings between them are frequent. The continuo may be played by two instrumentalists, one improvising the chordal realization, for example on the harpsichord, the organ, or the lute.

Rameau (1726) analyzed some Corellian cadences in his *Nouveau système*, which led, among other things, to the theorization of the cadential 6/4. Bergé (2015) discusses how cadences structure Corelli music. Most of them mark transitions between phrases, and sometimes are recurring elements across movements (Libby, 1973). Wintle (2017b, p.33) states that "Corelli's trio-sonata textures have been defined in terms of the polarization of two intertwining violins, moving predominantly in parallel thirds, which are set apart from their supporting bass line". Wintle (2017a, p.56) also analyzes the Corelli sonatas across their rhythms and their modulations, and discusses structure through repetitions and expansions/extensions of

tonal and "large rhythmic units". He further states (p.32) : "Corelli's oeuvre is founded upon a fairly limited number of musical figures, or models, which are capable of sustaining a considerable variety of modes of presentation."

Here we analyze formal structures, texture, and imitation patterns, trying to describe such Baroque writing techniques. The imitation patterns are repeated, possibly transposed, and possibly used in melodic sequences. We release a dataset on all the 38 slow movements of the 24 Corelli's Church Sonatas from Op. 1 and Op. 3. These slow movements are relatively short, allowing in-depth analyses, and showcase a variety of textures and writing techniques with a focus on patterns. They already account for 954 measures, making this sub-corpus a suitable option for an initial modeling of textures. We began with *.musicxml* files released by Hentschel et al. (2021) [1], built by curating data from Craig Sapp's KernScores [2] and the Pepusch (1740) edition of Corelli's Trio Sonatas Op. 1 and Op. 3 available from the International Music Score Library Project (IMSLP) [3]. We complement the annotations from Hentschel et al. (2021) by new annotations on texture and writing techniques. The fact that, including us, three different groups have already worked on and improved this material shows the advantage of such open-data research.

Modeling texture and imitations is less common than modeling harmonic functions and is open to debate. With our formalization, we aim to limit certain biases, as for example, those introduced by inconsistent or overly intuitive annotation. However, we acknowledge that any analysis of musical texture, a domain that is not yet well established, is inevitably shaped by our perspectives and interpretative choices, which we consider an inherent and acceptable part of the process. An example of this approach can be seen in the *Grave* of the first Trio Sonata from Op. 1 in F major (Figures 1 and 2).



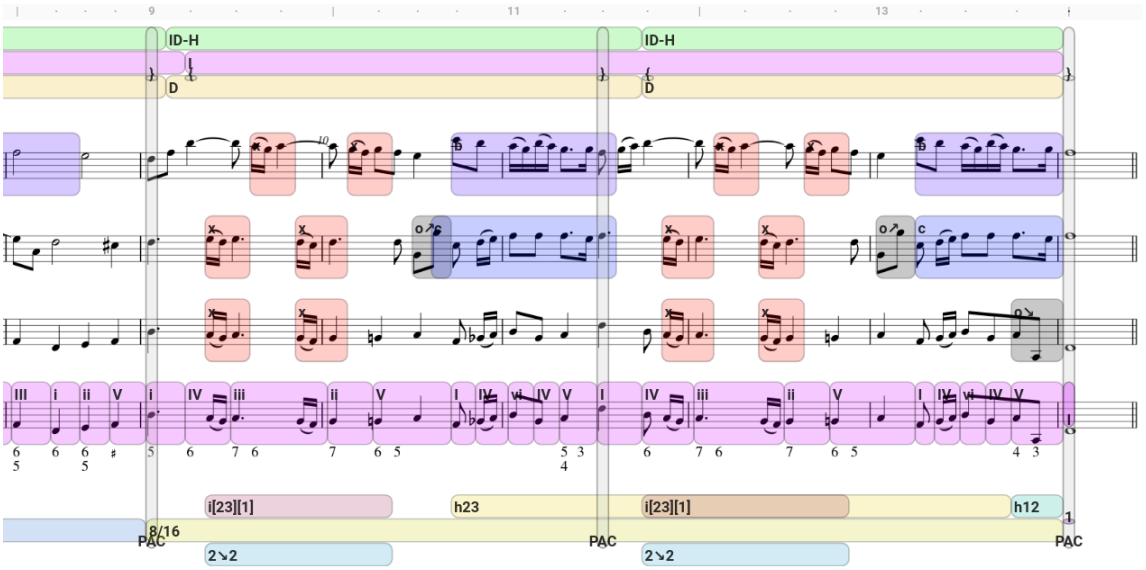


Figure 1. Analysis of the first movement (*Grave*) of the Trio Sonata in F major, Op. 1 No. 1, annotations as visualized on *Dezrann* : https://www.dezrann.net/~/corelli-trio-sonatas/op1.1.1. Patterns are annotated (***a***, ***b***, ***c*** | ***x*** (imitation)), and further labels are on phrase textures (Homorhythmy (***H***), Imitation with suspensions (***ID***)), pattern textures (homorhythmy (***h***), imitation (***i***))), rhythm density (from whole note (***1***) to sixteenth note (***16***)), melodic sequences (number of occurrences, direction, diatonic interval), ***o***↗, ***o***↘ (octave jumps) and on cadences (Perfect Authentic Cadence (***PAC***))

Bar	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Local key	I				V	I	vi		I					
Structure	A		B			C			D			D		
Cadence	PAC			PAC			PAC			PAC				PAC
Textures	H		H-ID			H			ID-H			ID-H		
	h12		i[1][23]						i[23][1]		h23	i[23][1]	h23	h12
Rythmic density	4/8	8/16	4/8	8/16		2/4			8/16					1
Suspension		43		76	76	98			76	76		43	76	76
														43
Melodic sequence				3↘22					2↘2			2↘2		
Violin I	b			x	x	x	b		x	x	b	x	x	b
Violin II	c			x	x	x			x	x	o↗ c	x	x	o↗ c
Continuo	a	o↘	a											o↘

Figure 2. Analysis of the first movement (*Grave*) of the Trio Sonata in F major, Op. 1 No. 1, using the same notations as in Figure 1.

The *Grave* contains five phrases, ***A, B, C, D, D***, the last phrase ***D*** being (almost) repeated. Phrase ***B*** modulates to C major and phrase ***D*** modulates to D minor. Phrases ***A*** and ***C*** are mostly homorhythmic (***H***), while phrases ***B*** and ***D*** have parts with imitation with suspensions (***ID***). The second phrase ***B***

combines some transformed material from ***A*** (in homorhythmy) with an imitation of the pattern ***x*** (3 occurrences in each voice) that yields a 7-6 suspension in bar 4. The annotations further specify which voices are involved and how they interact. In the first bar of the *Grave*, the two violins are in homorhythmy (***h12***). The imitation in ***B*** is annotated ***i[1][23]*** as the first violin (voice 1) begins, and then the other group, second violin and continuo (voices 2, 3), continues. In the third bar, during the imitations, our interpreted r*hythmic density* is annotated ***8/16***, as an "eighth note / sixteenth note" rhythmic pattern dominates this section. The thematic patterns ***a***, ***b***, and ***c*** are exposed in phrase ***A*** (and are reused and transformed in later phrases). A short imitation pattern ***x*** is repeated and transposed during melodic sequences. The melodic/harmonic sequence in ***B*** is annotated as ***3↘22***, that is three occurrences, descending each by a second (either major or minor).

METHODS

To build the dataset, we first segment and label phrases according to their similarity, using phrase boundaries and cadences from Hentschel et al. (2021) for their delimitation. The movements contain between 2 and 11 phrases (5.26 on average).

The central addition of this dataset is *texture* and *patterns*. These annotations take into account both the perception of the musical streams, and the written elements that are predominant in the analysis of the score. We identify patterns as musical elements repeated at least twice in the movement, which may arise from a paradigmatic analysis (Ruwet, 1972). Imitation patterns are usually short (3-5 notes), located inside melodic sequences. Patterns may occur in overlapping textural zones. Outside of imitations, one finds thematic patterns in homorhythmy. The same pattern can also be involved in both the end of an imitation and the beginning of the following homorhythmy. The patterns are mostly independent, except in a very few cases, which are too limited in number to be explicitly encoded in the annotation syntax.

Each phrase is thus annotated with its *texture* (Table 1), either imitation with or without suspension, or mostly homorhythmic (capital letters, ***H, I, ID***). What is considered "mostly" homorhythmic is subject to interpretation: Typically, at least 80% of the beats are synchronized between the involved voices. Some passages may still include slight variations, such as the sixteenth notes in the second bar of Violin II on the Figure 1. In the case where a phrase contains two textures without a clear predominance of one of them, we indicate two textures on a segment, as in the case of phrases ***B*** and ***D*** in Figure 1 which are respectively annotated ***H-ID*** and ***ID-H***, to render the order in which the textures occur.

We then delimited the homorhythmy and imitation zones (lowercase letters, ***h/i/a***, Table 2), indicating the involved voices (see ***h12*** on Figure 1). We start imitation zones from the beginning of the first pattern and stop them at the end of the last one. Homorhythmy is annotated even when the homorhythmy is not "perfect", as in the first measures of the Grave. Some homorhythmies are rather annotated as parallelism (***p***). Note that zones may overlap. Moreover, some homorhythmic zones (***h***) are annotated in non ***H*** phrases. Crossed imitations with exchange are rare (1.7.2 and 1.12.1)

Phrase name and variation (6 structure types, 6 variation types, 200 labels) | ***A***, ***B***, ***C***, ***D***, ***E***, ***F***
Phrases with small variations (as the last note on the cadence on C on Figure 2) are reported as the same phrase.
Phrases can be tagged with :

- \+ : mainly extension of the previous phrase
- \- : mainly truncation from its end
- ′ : another mutation

Phrase texture type (7 types, 200 labels) | ***ID*** (65), ***H*** (76), ***I*** (26), ***I-H*** (7), ***ID-H*** (9), ***H-ID*** (8), ***H-I*** (9)

- ***ID*** : Imitation with suspensions
- ***I*** : Imitation without suspension
- ***H*** : (Mostly) homorhythmic, no imitation

Melodic move and imitation texture (3 types, 24 sub-types, 328 labels) *(see Figure 3)*

- ***h*** : homorhythmy, some of them being ***p*** : parallelism
- ***i*** : imitation (contrasting groups, "call and response" about the same material in the two groups)

Rhythm density : main interpreted rhythmic figure (5 types, 15 sub-types, 376 labels)
From whole note (***1***) to sixteenth note (***16***). Can be combined (**2/8**). Hemiolas are denoted with ***hem***.

1 (37), ***1/2*** (51), ***1/4*** (25), ***1/8*** (6), ***1hem*** (9), ***2*** (11), ***2/16*** (7), ***2/4*** (69), ***2/8*** (54), ***2hem*** (3),
4 (3), ***4/16*** (30), ***4/8*** (53), ***4hem*** (1), ***8/16*** (17)

Table 1. Annotations on structure and on textures

Thematic pattern (5 types, 107 labels) | ***a***, ***b***, ***c***, ***d, e***

Imitation pattern (8 types, 610 labels) | ***x***, ***y***, ***z***, ***t***, ***u***, ***v***, ***w***, ***s***

Melodic/harmonic sequence (12 types, 2 directions, 6 number of times, 159 labels) | indicating number of occurrences (between 2 and 5), direction (either downwards↘ or upwards↗), and diatonic intervals *(see Table 4)*

Octave jump (391 labels) | ***o***↗ ascending (198), ***o***↘ descending (205) *(see Table 5)*

Table 2. Annotations on patterns, melodic sequences, and octave jumps. Note that there are never more than 5 thematic and 8 imitation patterns in a single movement of our corpus.

We complete the annotation of patterns by non-overlapping annotations for rhythmic densities, labeling the most represented rhythmic figures in each zone, the hemiolas, the melodic sequences, and octave jumps (Table 2). The dataset is completed with harmonic data from Hentschel et al. (2021) (Table 3).

Cadence (5 types, 237 labels) | ***DC*** (2), ***EC*** (8), ***HC*** (53), ***IAC*** (22), ***PAC*** (152)

Local key (11 types, 160 labels) | such as ***I*** or ***V***

Chord function/degree (17 types, 2806 labels) | such as ***I*** or ***vi***

Chord quality (1297 labels) | such as ***5***, ***64***, ***7*** or ***9***

Phrase borders (235 labels)

Table 3. Annotations on structure and harmony from Hentschel et al. (2021)

DATA AVAILABILITY

The dataset contains over 7,000 labels. The annotations were proposed by the first author and curated by all authors. We also provide bar-level synchronization, achieved on the *Dezrann* platform (Garczynski, 2022), for a recording by the London Baroque orchestra directed by Charles Medlam (Harmonia Mundi, 2007). Currently, we are unaware of any complete open-source recordings of these sonatas. However, we intend to publish synchronization data for them when they become available.

The dataset, comprising scores, annotations, synchronization data, and metadata, is accessible via the open license CC-BY-NC-SA 4.0 on a git repository accessible from algomus.fr/data and on the long-term institutional repository https://doi.org/10.57745/QUZ8KO. It is also accessible on the *Dezrann* platform (dezrann.net/explore/corelli-trio-sonatas), featuring synchronized playback with the recording.

DESCRIPTIVE STATISTICS

Annotating textures helps us to understand the construction of Corelli's sonatas. The instruments play on average 41% in homorhythmy (including 8% in parallelism) and 43% in imitation (Figure 3). The vast majority of imitations occur between the two violins, and thus the continuo rarely plays in imitation (less than 5%). Conversely, the continuo frequently plays in homorhythmy with the violins, though a bit less with the first violin.

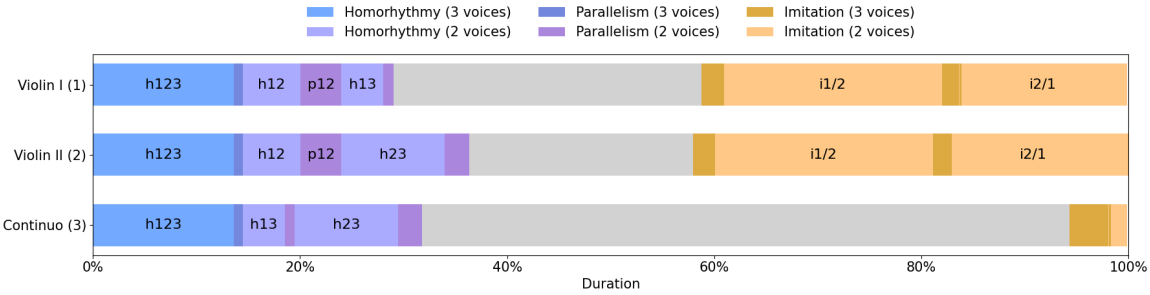


Figure 3. Homorhythmy, parallelism, and imitation across the three parts

	Number of repetitions						Number of intervals between occurrences								
	1.5	2	2.5	3	4+	TOTAL	P1	m2 M2	m3 M3	P4	P5	m6 M6	m7 M7	TOTAL	
↗	3	**59**	1	3	2	68	21	28	10	17	9	6	2	72	21
↘	4	**48**	1	**24**	7	84		**79**	12	14	15	2	0	121	
↗↘/↘↗	0	0	1	5	1	7									
TOTAL	7	107	3	32	10	159	21	107	22	31	24	8	2	215	

Table 4. Melodic sequences according to their direction, number of repetitions, and intervals

Most of the melodic sequences have two occurrences (ascending or descending ones) or three occurrences (mostly descending ones). Most of the intervals in those melodic sequences are descending minor and major seconds (Table 4). Octave jumps are mostly (69%) located on the bass, with 64% of them being descending ones, often just before the cadence, as in bars 2 and 13 of the *Grave*. Conversely, 79% of the octave jumps on the violins are ascending (Table 5). The majority of hemiolas are positioned at the end of the movements (Figure 4), creating a slowing-down effect.

	Violin I	Violin II	Continuo	TOTAL
↗	42	57	99	198
↘	17	9	179	205
TOTAL	59	66	278	403

Table 5. Ascending (***o***↗) and descending (***o***↘) octave jumps on the three parts

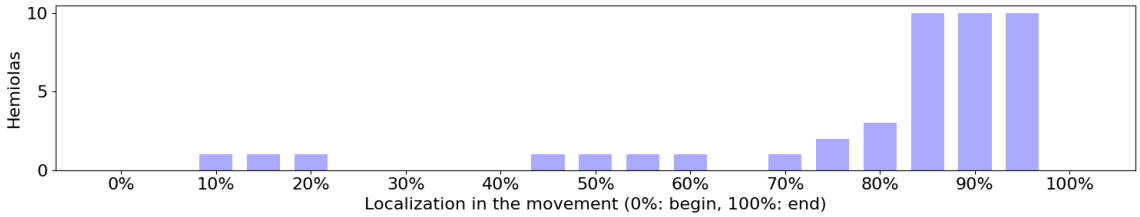


Figure 4. Localization of the hemiolas

PERSPECTIVES

Overall, this dataset offers a more abstract and high-level view of musical structure than Hentschel et al., especially by annotating common textures in the Baroque style, imitation patterns, and melodic sequences. Defining texture is complex, and there is no established consensus on how to annotate it. Here, we proposed a consistent method to describe these textures with greater precision. We believe that this annotation data can be valuable for systematic musicology and computer music, benefiting people working not only on Baroque music but also on other styles. Indeed, imitation textures and other high-level musical structures occur across various genres, making this dataset a useful resource for projects in music analysis and music generation. For instance, being able to model a "slow movement, where the second phrase features imitation with a specific pattern and ends on a half cadence" can support both human and automated analysis, as well as generative approaches, whether through procedural algorithms or machine learning.

More broadly, this initiative represents a step toward formally modeling and discussing musical texture, providing a framework for understanding this essential aspect of music. However, as the dataset is focused on these slow movements, which are relatively homogeneous, the rhythmic variety and contrast may be here limited. Further studies could target fast movements in Corelli's sonatas and other works to model textures more comprehensively.

NOTES

[1] https://dcmlab.github.io/corelli/
[2] http://kern.humdrum.org/search?s=t&keyword=Corelli
[3] https://imslp.org/wiki/Category:Corelli,_Arcangelo

REFERENCES

Allsop, P. (1992). The Italian 'Trio' Sonata: From Its Origins until Corelli. Oxford Monographs on Music.

Bergé, P., & Neuwirth, M. (2015). What is a Cadence?: Theoretical and analytical perspectives on cadences in the classical repertoire. Leuven University Press. https://doi.org/10.2307/j.ctt14jxt45

Garczynski, L., Giraud, M., Leguy, E., & Rigaux, P. (2022). Modeling and editing cross-modal synchronization on a label web canvas. In *Music encoding conference* (MEC 2022). https://doi.org/10.17613/krxd-pq66

Hentschel, J., Moss, F. C., Neuwirth, M., & Rohrmeier, M. (2021). A semi-automated workflow paradigm for the distributed creation and curation of expert annotations. In *International Society for Music Information Retrieval Conference* (ISMIR 2021). https://doi.org/10.5281/zenodo.5624417

Libby, D. (1973). Interrelationships in Corelli. *Journal of the American Musicological Society*, *26*(2), 263-287. https://doi.org/10.2307/830667

Pepusch, J.C. (1740). The score of the four setts of sonatas compos'd by Arcangelo Corelli, for two violins & a bass. *Ed. Benjamin Cooke, London, 1st éd. Thomas Baker Collection at the University of Western Ontario Music Library https://opac.rism.info/rism/Record/rism990011184*

Pincherle, M. (1954). Corelli et son temps. Edition Le Bon Plaisir / Plon.

Rameau, J. P. (1726). *Nouveau système de musique théorique*. Ballard. Chap. 23, 94-106.

Rothstein, W. (2006). *Transformations of Cadential Formulae in the Music of Corelli and his Successors*, Third International Schenker Symposium.

Ruwet, N. (1972). *Langage, musique, poésie*, Paris, Éditions du Seuil, 100-134.

Sanna, A. (2013). Contrapuntally crafted, harmonically eloquent. Corelli's sonatas and the compositional process in the late 17th century. *Basler Jahrbuch für Historische Musikpraxis*, *37*, 73-88.

Wintle, C. (2017a). Corelli's Rhythmic Models: Dance Movements in the Seventeenth and Eighteenth Centuries. *Intégral*, *31*, 51-61.

Wintle, C. (2017b). Corelli's Tonal Models: The Trio Sonata Op. 3, No. 1. *Intégral*, *31*, 31-49.