

Streaming complexity in the Renaissance Mass Ordinary cycle

Finn Upham^{1†}

Julie Cumming²

¹ RITMO, Department of Musicology, University of Oslo, Oslo, Norway

² Schulich School of Music, McGill University, Montreal, QC, Canada

† Corresponding author: finn.upham@gmail.com

Published 16 December 2021; <https://doi.org/10.18061/FDMC.2021.0013>

Author video presentation and/or other conference material: <https://doi.org/10.17605/OSF.IO/8D45J>

Abstract

Complexity of a piece should also be related to its intended use and context. The Mass Ordinary cycles of 15th- and 16th-c. Europe were works of exceptional formal consistency, and if complexity is relevant to the role of a Kyrie, Gloria, or Agnus Dei, this should be measurable from the scores of surviving examples. We evaluate an aspect of complexity generalizable to past church goers from the consequences of low-level auditory streaming principles. The streaming complexity estimate presumes that this music tends to be heard as an integrated stream within the rich scene of the church service unless the voices diverge through separation cues such as independent onsets and contrary motion. In a corpus of more than 200 mass cycles composed over 150 years, we find significant differences in average streaming complexity over time and between movements. The Agnus Dei tends to be more complex, the Credo, with a long text, is usually low, depending on the composer. Results suggest streaming complexity may indeed be tuned for the intended use of a work. By bringing musicological evidence into corpus analysis, we can consider this feature in relation to the role of music beyond present day patterns of consumption.

KEYWORDS: *auditory streaming, Renaissance mass ordinary cycles, complexity, corpus studies*

Introduction

Complex polyphony bloomed in Renaissance church music as the mass ordinary cycle emerged in the early 15th century. Was variation in this complexity a form-defining feature that could have been followed by those attending services in the 15th and 16th centuries? Without sound recordings or detailed reports by contemporary listeners, we can only approximate what could have been perceived from symbolic representations of these scores and records of listening conditions. The auditory streaming complexity estimate is calculated as a continuous feature reporting a moment-by-moment tally of how many independent streams or objects might be heard by a casual listener, without explicitly trying to count. This estimate evaluates how individual notes contribute to the blending or separation of each part from the rest. While some cues for sound stream

merging and segregation have very robust effects on the attentive listener (Bregman, 1994), attention also plays an important role in how they interpret the auditory scene (Cusack et al., 2004). Perception of sound sources comes from interactions between bottom-up sensory processing and top-down expectations tuned to the signals detected and to the orientation of the listener's auditory focus (Shinn-Cunningham et al., 2017). The exact perceptions of a 15th-century European church goer cannot be fully modelled today, but we expect that much of their low-level sensory processing to be similar to those of participants in 20th- and 21st-century psychoacoustic experiments with moderate attentiveness.

The principles of auditory scene analysis have been applied to counterpoint and polyphony before. David Huron's work on voice leading identified many connections between the cues influencing sound stream integrity and the rules of species counterpoint (Huron, 2001). The relationship between streaming cues and the merging of parts in polyphonic works has also been looked at experimentally with perceptual judgments. Ben Duane asked participants to judge the audible merging of parts in string quartet excerpts (played in midi) and compared these perceptual accounts to quantifications of stream separation cues discerned from the scores (2013). Duane's results suggested asynchronous note onsets and harmonic dissonance were important factors decreasing the merging of two string parts into a blended stream. Emilios Cambouropoulos combined similar features to identify voices within polyphonic piano music, using onset synchrony, pitch comodulation, and stream continuity to calculate which notes would be integrated vertically and horizontally (2008). This project applies these principles to a corpus of Renaissance mass ordinary cycles to evaluate whether this complexity varies meaningfully with form.

The polyphonic mass ordinary cycle in the Renaissance is an ideal genre for a computational study of streaming complexity. The mass ordinary cycle is a



polyphonic setting of five texts that were said or sung at every Catholic mass, and each of the mass ordinary texts came into the mass at a different time and served a different liturgical function. It is the largest and most complex musical genre in the Renaissance, and there are many surviving mass cycles. To a 15th-century churchgoer, this polyphonic music sounded elaborate, festive, and special. The alternation between the complex music of the mass ordinary cycle (where each movement begins with similar music) and the chant and spoken text of the rest of the mass could have provided an overall unity to the ceremony. The standardized sequence of the five movements, each with a different approach to text-setting and structure, might have allowed listeners to locate themselves in the ceremonial time of the mass, even if they could not follow what was being said. The melismatic Kyrie and Agnus Dei, both of which have three sections, serve as a kind of internal frame for the mass as a whole. The long syllabic Gloria and Credo movements articulate the liturgy of the word. The melismatic Sanctus and Agnus frame the mystery of transubstantiation. The music of the polyphonic mass ordinary would have provided an aural focus for parishioners, for being “present with the body and intent with the mind” when these services were not expected to be fully heard and understood by all attending.

This project brings these concepts of separation and merging cues to calculate streaming complexity on a large symbolic corpus of polyphonic vocal music and evaluates whether this feature of polyphonic writing changes substantively over time and in ways that might have been relevant to the composers and audiences of the day.

Method

Dataset of symbolic music files

The symbolic music files for this corpus analysis came from two public collections of polyphonic Renaissance music: The Josquin Research Project (JRP), an open access database of Renaissance scores in multiple symbolic formats (Rodin & Sapp, 2019) and RenComp7, a repository of MIDI scores with additional polyphonic vocal works of the Renaissance (McKay et al., 2017). The RenComp7 dataset provided later works: 18 mass ordinary cycles by Tomás Luis de Victoria and the 101 by Giovanni Pierluigi da Palestrina. Another 84 mass ordinary cycles were downloaded from JRP in kern format, works by 13 identified composers and a few without clear attribution. In total, the corpus includes 216 masses by at least 15 different composers.

Using available musicological evidence, these works were additionally grouped by estimated time of composition to the quarter century.

Nearly all of these masses included the typical five movements of the mass ordinary: Kyrie, Gloria, Credo, Sanctus, and Agnus Dei. A few were accompanied by mass proper movements such as the Introit, and these individual movements were discarded prior to analysis. Mass cycles from later periods often contained six or seven or eight movements if the ordinary movements were broken into first and second parts (Agnus Dei), or the Benedictus separated from the first part of the Sanctus. For this analysis, these components were concatenated to match the standard five movements, all sharing the same text and liturgical role. A few mass cycles were incomplete because of missing movements or files that did not render properly; the remainder of these cycles were retained for analysis. Across all 216 masses, 1042 movements were studied, each containing 3 to 12 voice parts.

Calculating Auditory Streaming Complexity

Unison chant was a common part of church music at this time, in which coordination would favour hearing a group of singers as a single sound source. Polyphonic music emerged in contrast to this aural unity, with an emphasis on counterpoint resulting from the independence of voices in pitch, rhythm, and text. With a combination of Music21, SciPy, Numpy, and Pandas (Oliphant, 2007; Cuthbert, 2019; McKinney, 2010), we evaluated in every score three types of sound events that can contribute to the separation or merging of voices:

1. Voice Entries: when a part enters for the first time or after a rest.
2. Pitch Changes: when a part changes pitch and in which direction.
3. Note Onsets: the beginning of each articulated note in the score.

These are then reduced by their alignment with the same events in other (higher) parts, and scaled by importance. We treat voice entries as stronger separation cues than pitch changes, which are stronger than note onsets in the sustained singing style of this genre.

Auditory streaming is cumulative and a bit sticky: once sounds in the auditory field have been attributed to separate sources, we try to hold on to that organisation until contradictory cues overwhelm it (Bregman, 1994). Because the ear rapidly reassess soundscapes (Shinn-Cunningham et al., 2017), the effect of separation cues per part were integrated with a short linear decay,

reducing at a rate of 0.2 per semi-minim or quarter note. These calculations translate to the separation effect of a solo voice entry (against other sound parts) dissipating within 1-1.5 seconds depending on performance tempo.

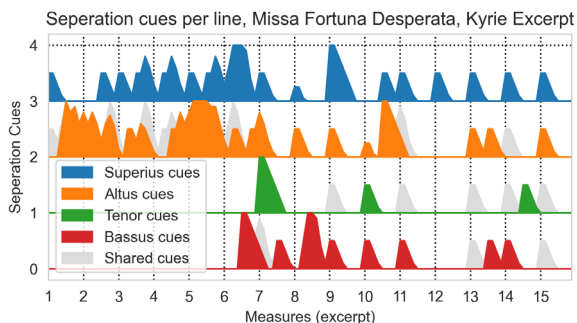


Figure 1: Integrated stream separation cues per part for a 16-m excerpt of a mass by Josquin de Prez. In grey are cues shared with higher voices, contributing to merging of parts, in colour are the integrated separation cues.

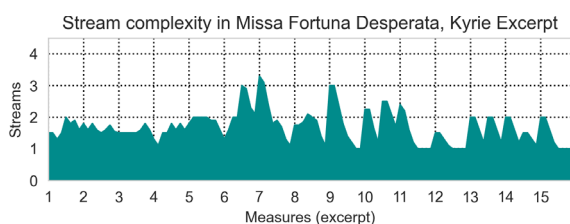


Figure 2: Auditory streaming complexity estimate for the 16-m excerpt reported in Figure 1. The opening duo (1-5 mm) maintains strongly separated voices. In mm 12-15, four parts are highly homorhythmic and reduced.

For the corpus analysis, the feature of streaming complexity was averaged per movement and the relative mean streaming complexity calculated by subtracting the average mean per mass cycle. These summary statistics are better interpreted as the center of mass for streaming complexity than representative of the most common values.

Results

Streaming complexity over time

The first result demonstrates that streaming complexity was not a static quality in renaissance polyphony. Linear regression on random 50 movement subsets from each quarter century era showed a significant increase in estimate values over time (Figure 3). Era significantly predicted mean streaming complexity, $b = .12$, $t(258) = 17.3$, $p < .001$. Era explained a significant portion of variance between mass cycles, $R^2 = .55$, $F(4, 245) = 79.9$, $p < .001$.

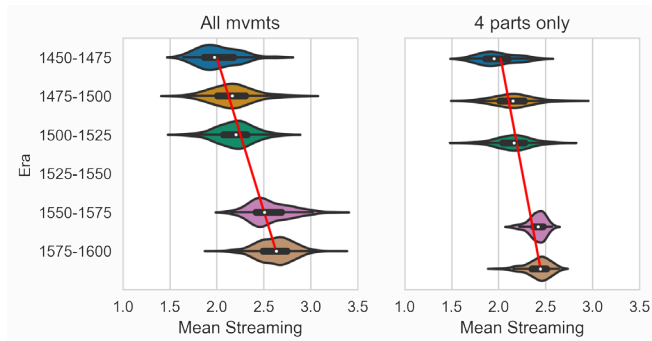


Figure 3: Increasing streaming complexity over time. The distribution of mean streaming complexity per movement increases substantially over successive eras. The effect is substantial when all movements are considered and when considering only movements with 4 voice parts. Regression line in red.

It is known that later works tended to include more parts, which makes higher streaming complexity values possible. However, the increase cannot be attributed only to additional voices, as the effect is noticeable across works with the same number of parts. The 4-part movements shown in the second plot of Figure 3 shows a comparably significant trend: $b = .088$, $t(258) = 15.2$, $p < .001$, $R^2 = .48$, $F(4, 245) = 63.2$, $p < .001$.

Differences between movements

While masses changed in streaming complexity over time, the hierarchy between movements appears to be fairly robust. A two-factor ANOVA (movement and composer) on relative mean streaming values showed a significant effect of movement and a significant interaction between movement and composer for both statistics (Vallat, 2018). The main effect of movement yielded an F ratio of $F(4,788) = 88.1$, $p < .001$, indicating

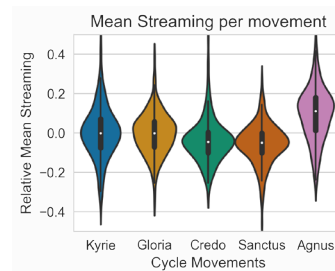


Figure 4: Distributions of relative complexity values per Mass Ordinary Cycle movement. Violin plots report the median, quartiles, and smoothed distribution.

a significant difference in relative mean streaming complexity between movements. Pairwise differences between most movements were strong ($p < .001$), except for Credo-Gloria and Gloria-Sanctus ($p < .01$) and no significant difference between Credo-Sanctus and Gloria-Kyrie (Games-Howell post hoc test, allowing for unequal variance). The significant interaction between composer and movement, $F(64,788) = 3.4$, $p < .001$, suggests analysis of individual composers' patterns is warranted.

Most composers followed this softened V shape, shown most clearly in Palestrina's works (Figure 5). The one well-represented composer to differ substantially was Josquin des Prez. Most mass cycles confidently attributed to Josquin reaching relatively high streaming complexity in the Credo, a distinctive strategy.

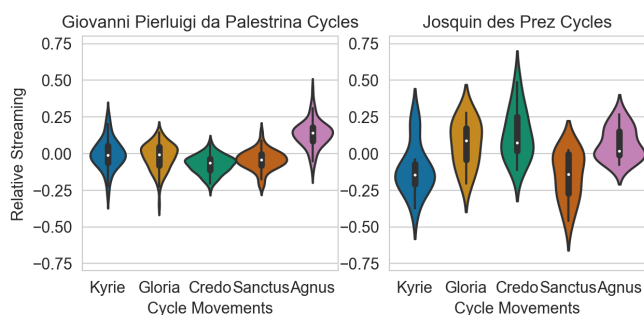


Figure 5: Movement trends for streaming complexity in masses by Palestrina and Josquin in violin plots.

Discussion

Masses composed between 1450 and 1600 show a trend of increasing complexity. Streaming Complexity Estimate averages fit a strong regression to the five quarter-century sets in the corpus. And the change was not only a consequence of more parts: a parallel analysis of movements with exactly 4 parts shows a similarly strong increase in streaming complexity. These findings correspond to what musicologists and performers already know about style change in the Renaissance, and in the mass in particular. Four-voice writing was typical before 1540, followed by more five-voice writing and experimentation with large numbers of voices in the late 16th century. Not only do we have more voices in the score, we now know that this increase in streaming complexity means masses were composed in a way that individuated voices from each other more of the time: less merging, more separation.

The relative streaming complexity of the different movements also corresponds to musicologists'

understanding of the conventions associated with each movement (Fallows, 2012). The subtle “V” shape indicates that streaming complexity decreases from the Kyrie to the Credo, and then ascends through the Sanctus to the highest complexity in the Agnus Dei. The Credo is normally the lowest in streaming complexity; it has the longest text (163 words), and tends to use syllabic text setting and homorhythm. The Sanctus, in contrast, is very melismatic; it has five sections (Sanctus-Pleni-Hosanna-Benedictus-Hosanna), and the Pleni and the Benedictus are normally written for a smaller number (2-3) of voices. It is sung during the liturgical high point of the mass, transubstantiation. The Agnus normally adds voices in the last of its three sections and often features special effects: canonic structures, a cantus firmus in very long notes, or wider ranges (Fallows, 2012). The Agnus is designed to go out with a bang. Overall, the consistency of the “V” shape in our corpus of mass cycles reinforces our initial claim that the sung mass ordinary cycle served as a way to mark experiential time in the context of the larger mass liturgy, perceptible without needing to understand the words.

As in any musical genre, composers could make different choices about which movements should be most complex. Within this corpus only a few composers were represented by enough works to test quantitatively, but the one whose masses stood out most clearly was Josquin des Prez. Unlike the general average hierarchy in the masses of Palestrina, Victoria, and de la Rue, the seventeen masses that are reliably attributed to Josquin have streaming complexity averages that put the Credo movement high in complexity and the Sanctus relatively low. This inversion of the normal hierarchy raises questions about whether those influenced by his work would have followed suit, but more scores are needed for such a study.

Conclusion

The Auditory Streaming Complexity Estimate was constructed out of principles of auditory perception without optimisation of parameters to fit reported perception. But even without the opportunity to sharpen the model, this measure appears to demonstrate meaningful patterns in this corpus of Renaissance masses.

Over 150 years, this liturgical genre increased in the average streaming complexity, as calculated from the interactions among the voices, and beyond the effect of the number of voices. The five movements of the mass, with their own liturgical functions and formal

contributions to the mass service, tended to differ in their streaming complexity, creating a hierarchy where some movements tended to be more complex than others across the repertory, according to either feature. These differences were not absolute, but they were still common enough for multiple composers to show the same tendency, a soft V shape with the central text-heavy Credo maintaining lower streaming complexity and the closing glorifying Agnus Dei reaching the greatest heights. This persistent form allows us to imagine how the cycle could have functioned to mark time for the Renaissance listener. The exception, Josquin des Prez, appears to have had his own rules for which movements were most complex in terms of auditory streams.

This estimate of auditory streaming complexity could be adapted to include more information-related stream formation. Motivic repetition, for example, sets up expectations. This corpus covers the period in which imitative texture became dominant for sacred music, and it may well play a role in the growth in complexity already shown. Another research question opened by this analysis is whether composer-specific trends carried forward in ways consistent with records of teaching and region.

It would be immensely useful to tune the Streaming Complexity Estimate directly to reported or measured experiences of listeners. But even without this stage of validation, the strength of the results in this corpus analysis suggest that it establishes the importance of texture as a musical feature that changes within a multi-movement work, and over time. Our calculations were tailored to this genre of music, to specific types of listeners (non-expert), and to a particular listening context. Similar evaluations could be performed on other genres of polyphonic music, such as motets and madrigals, and with some additional changes, to classical genres such as the string quartet.

Acknowledgements

Thanks to Ichiro Fujinaga and the SIMSSA team for their contributions and expertise. Financial support was generously provided by the granting agencies FRQSC (Fonds de recherche du Québec – Société et culture) and SSHRC (Social Sciences and Humanities Research Council of Canada).

References

Bregman, A. S. (1990). *Auditory Scene Analysis: The Perceptual Organisation of Sound*. MIT Press.

- <https://doi.org/10.7551/mitpress/1486.001.0001>
- Cambouropoulos, E. (2008). Voice and stream: Perceptual and computational modeling of voice separation. *Music Perception*, 26(1), 75–94. <https://doi.org/10.1525/mp.2008.26.1.75>
- Cusack, R., Deeks, J., Aikman, G., & Carlyon, R. P. (2004). Effects of location, frequency region, and time course of selective attention on auditory scene analysis. *Journal of Experimental Psychology: Human Perception and Performance*, 30(4), 643. <https://doi.org/10.1037/0096-1523.30.4.643>
- Cuthbert, M. S. (2019). *Music21 Toolkit* (Version 5) [Python]. cuthbertLab. <https://web.mit.edu/music21/>
- Duane, B. (2013). “Auditory Streaming Cues in Eighteenth- and Early Nineteenth-Century String Quartets: A Corpus-Based Study.” *Music Perception: An Interdisciplinary Journal* 31(1), 46–58. <https://doi.org/10.1525/mp.2013.31.1.46>
- Fallows, D. (2012). The last Agnus Dei: Or: The cyclic Mass, 1450-1600, as forme fixe. In A. Ammendola, D. Glowotz, & J. Heidrich (eds.), *Polyphone Messen im 15. Und 16. Jahrhundert: Funktion, Kontext, Symbol* (pp. 53–64). V & R unipress.
- Huron, D. (2001). Tone and Voice: A Derivation of the Rules of Voice-Leading from Perceptual Principles. *Music Perception: An Interdisciplinary Journal* 19(1), 1–64. <https://doi.org/10.1525/mp.2001.19.1.1>
- McKay, C., Cumming, J., and Fujinaga, I. (2017). Characterizing Composers Using JSymbolic2 Features. *Extended Abstracts for the Late-Breaking Demo Session of the 18th International Society for Music Information Retrieval Conference*. Suzhou, China.
- McKinney, W. (2010). Data structures for statistical computing in Python. In S. van der Walt & J. Millman (Eds.), *Proceedings of the 9th Python in Science Conference* (pp. 51–56). <https://doi.org/10.25080/Majora-92bf1922-00a>
- Oliphant, T. E. (2007). Python for scientific computing. *Computing in Science & Engineering*, 9(3), 10-20. <https://doi.org/10.1109/MCSE.2007.58>
- Rodin, J., Sapp, C. The Josquin Research Project. Accessed December 12, 2019. <http://josquin.stanford.edu/>.
- Vallat, R. (2018). Pingouin: Statistics in Python. *The Journal of Open Source Software*, 3(31), 1026. <https://doi.org/10.21105/joss.01026>