Annotation of Medieval Music Facsimiles Using ‘Good Enough’ OMR

Joshua Stutter
University of Glasgow, UK
j.stutter.1@research.gla.ac.uk

Abstract

The *Clausula Archive of the Notre Dame Repertory* (*CANDR*) is an in-progress PhD project with the aim of cataloguing, transcribing and analysing digital facsimiles of the thirteenth-century repertory commonly termed *Notre Dame polyphony*, and a secondary aim of providing new datasets and analytical tools for studying medieval polyphony. This poster highlights the use in the project of (a) a new methodology for de-skewing facsimile images, and (b) average symbol masks in an OMR–enhanced workflow with an emphasis on creating an OMR workflow that is ‘good enough’ to accelerate the annotation of an image dataset of particularly transitional notation.

1 Introduction

As far as medieval repertories go, it is difficult to locate any that is more significant to the development of ‘Western music’ than the Notre Dame (ND) school, a thirteenth-century collection of hundreds of pieces of polyphony on Gregorian chant. These settings of polyphony were likely initially intended for performance during services at Notre Dame de Paris, but it is now recognised that this particular style of polyphony spread throughout Western Europe in the mid to late thirteenth century [4].

This repertory is vitally important to the understanding of the nature of performance, rite, notation and the concept of ‘composition‘ in the medieval period. However, those that do not study it in any detail may have only heard of this repertory historiographically: an important juncture in the history of the ‘Western canon’, where the quantum leap from monophony to polyphony, and from oral to literate was made, but often only seen as a fleeting reference before moving onto more compositionally stable ground.

A picture of the ND repertory as a period of medieval musical revolution of composers over improvisers is obviously a simplified one, but it is a view frequently taken even by recently–written history and reference books. For example, Edward Roesner’s article for *Grove* characterises ND polyphony as “composition in the modern sense” [9, p. 202–203]. The ND school therefore becomes a useful jumping-off point to introduce the concept of written rather than oral practice, as well as the idea of named composers rather than anonymous improvisations. As a result, study of the ND school has been burdened with the anachronistic requirement to consider the ramifications of the first ‘composers’, ‘compositions’ and written process in Western music, responsibilities somewhat undeserving of a repertory which is continually being shown to be influenced by oral processes [1, 13].

2 Research Context

Rather than laying the blame for the elevation of the ND school as the ‘crucible’ of Western art music squarely at the feet of music historians, the construction of the ND repertory to be a progression towards classical music in fact emerges from the field’s overreliance on the ideas of Friedrich Ludwig, concerning the creation and dissemination of the ND repertory. It is almost impossible to consider the topic of Notre Dame polyphony without Ludwig’s *Repertorium* being mentioned [7].

Now over a century old, Ludwig’s monumental work on the ND repertory dominates current thought as the key text to understanding the creation and dissemination of the repertory in manuscript. *Repertorium* catalogues the settings of polyphony in the central manuscripts and draws concordances between settings and between the genres of *organum*, *clausula* and *motet*, purely through textual apparata and an array of sigla geared towards canonicalisation.

Although *Repertorium* is a striking piece of scholarship that has deservedly stood the test of time for its attention to detail and scrutability, Anna Maria Busse Berger, among others, has criticised Ludwig's biases, adherence to written composition and tendency towards *Urtext* [2]. Busse Berger rightly finds fault in Ludwig's crucial role in constructing the repertory from an unknown set of jumbled manuscripts and turning it into the missing link in the 'progression' from monophony and improvisation to polyphony and written composition, which would in Ludwig's view ultimately culminate in Palestrina. Although Ludwig's biases are now well known, Ludwig's conception of the ND repertory as fixed, written composition pervades scholarship in this area and, perhaps unsurprisingly, his flawed ideas are still often taken as central tenets of the repertory. Work must be done to unpick Ludwig's ideas as objectively as possible and re-evaluate the ND repertory outside of the modern Western canon.

3 *CANDR*

The *Clausula Archive of the Notre Dame Repertor*y (*CANDR*) aims to deconstruct Ludwig's concept of the ND repertory and begin the process of concordance once again from the source manuscripts as they appear in facsimile. *CANDR*'s aim is to untangle some of these interrelationships not as Ludwig did by hand and sheer force of repertorial knowledge, but by quantifiable methods.

The central manuscripts of the ND repertory, F (476 folios), W1 (214 folios), and W2 (253 folios), have been digitised and released online under permissive licences by their host institutions. This project takes advantage of these facsimile images as the base from which to generate a suitable dataset for calculating concordances. In order to extract the notational content from these images and analyse their interrelationships, the project uses a bespoke tool for automatic transcription from tagged annotations of staves of music. This proceeds in three phases:

1. Creation of an online database and annotation-transcription tool to aid the input of a dataset of thirteenth-century polyphony (complete).
2. Input of the dataset using these tools (ongoing).
3. Design and creation of tools to analyse this dataset dynamically (not yet begun).

Most of these tools had to be created particularly for this project, as many preexisting tools have no architectural support for the extremely contextual, transitional and inherently pluralistic notation of the ND school [11], which depends on performance practice, written and oral transmission, scribal practice, and was neither written for nor is suitable for performance. In creating a more flexible framework for ND notation, execution of these first two stages have presented two broad issues that have had to be solved. Firstly, the three-dimensional irregularity of the facsimile images caused difficulties during the annotation process, causing the automatic transcription to be inaccurate and affine de-skewing to be unsatisfactory. Secondly, the effort that would be required to transcribe the manuscripts (943 folios total) by hand was not feasible in any reasonable period of time.

4 Regularisation

By far the two most common issues in the annotation of medieval facsimiles are (a) curved parchment from the binding not allowing the folio to rest flat during photographing; and (b) perspective effects caused by the camera not being placed directly above the surface. These issues make it impossible to, for example, trace a straight, horizontal staff line across the staff as the line may be curved, and one side of the staff may appear larger than the other as it is marginally closer to the lens.

Previous work in this area takes two divergent paths: either to attempt to de-skew staves using shearing and other affine transforms on the facsimile image [5], or to take the image as-is, and consider skewed elements such as staff lines to be a property of the image, electing to model them as smoothed and reconstructed curves [12]. *CANDR* takes a new approach, using a perspective transform in multiple rounds to de-skew images to yield straightened staves. To be precise, 'de-skew' is not the correct terminology here, as 'skew' relates particularly to two-dimensional affine transformations (translation, scale, rotation, shear). The perspective transform used in

CANDR is a three-dimensional projective transformation, capable of not only transforming any parallelogram, but any arbitrary quadrilateral.

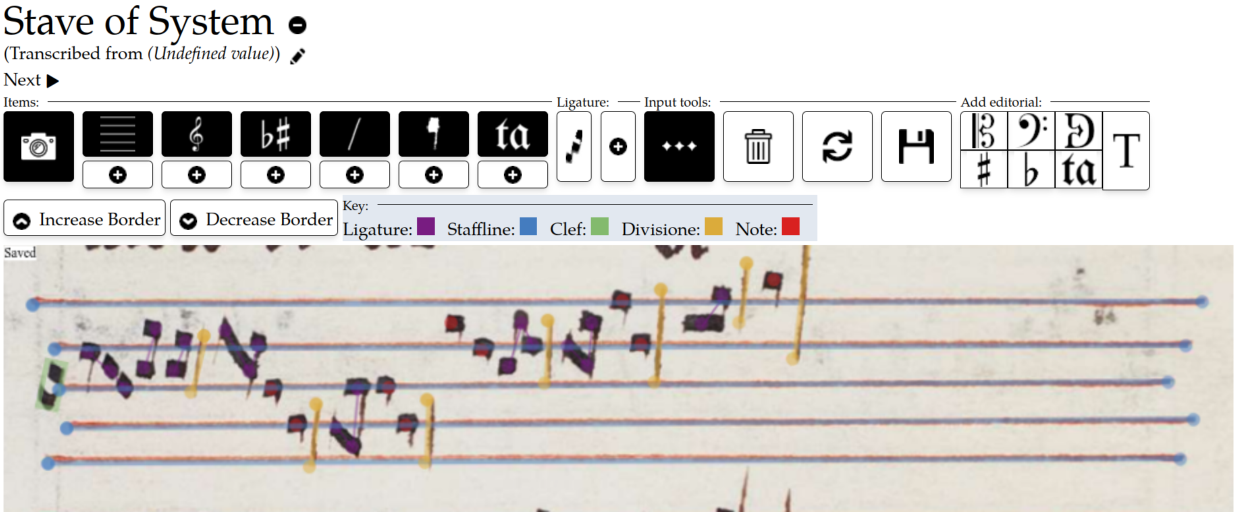


Figure 1: A fully-annotated stave: horizontal stafflines (dark blue), C clef (light blue), vertical *divisiones* (yellow), notes and ligatures (orange and red respectively).

First, the bounding quadrilateral of a system is manually annotated on a facsimile image, and an inverse perspective transform matrix is calculated that can transform that quadrilateral into a straightened rectangle. A second bounding quadrilateral for each stave on that system can be manually annotated upon the straightened system, and the same process applied to create individual straightened and cropped stave images. This process is integrated into the annotation workflow of *CANDR* and is calculated on-the-fly: annotating a system or stave creates a new object in the database along with its bounding quadrilateral. Viewing a system or stave in the visual editor is to view that item transformed into a straight rectangle using this transform. This method (a) mitigates the effects of parchment bend significantly, and (b) can eliminate not only rotation and skew but also perspective effects in the facsimile. There is also the added advantage of presenting to the user a clear and straightened image.

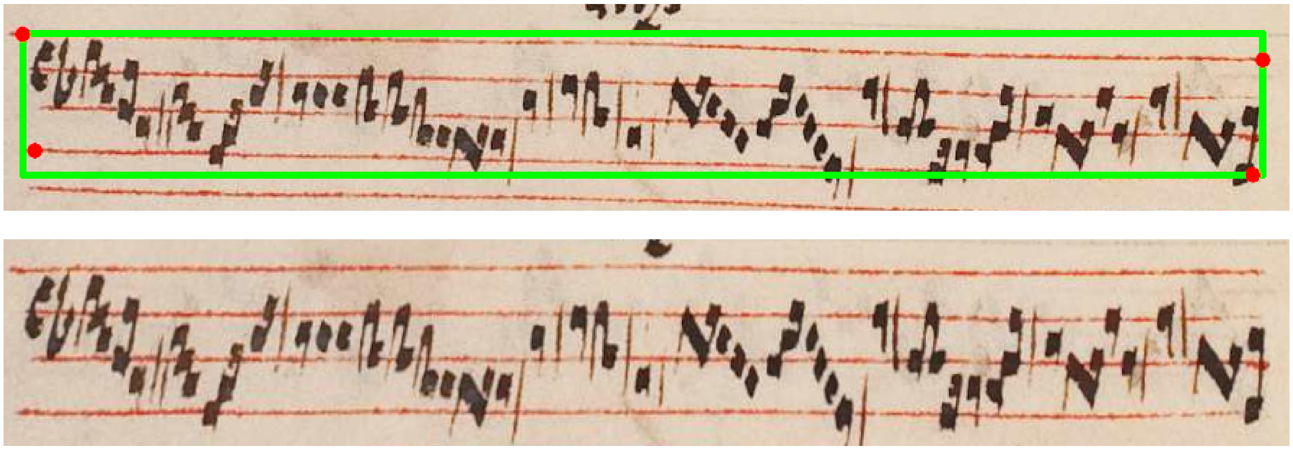


Figure 2: A real-world example: Before transform of four–line staff with staff boundaries and target rectangle marked (top), and after transform (bottom).

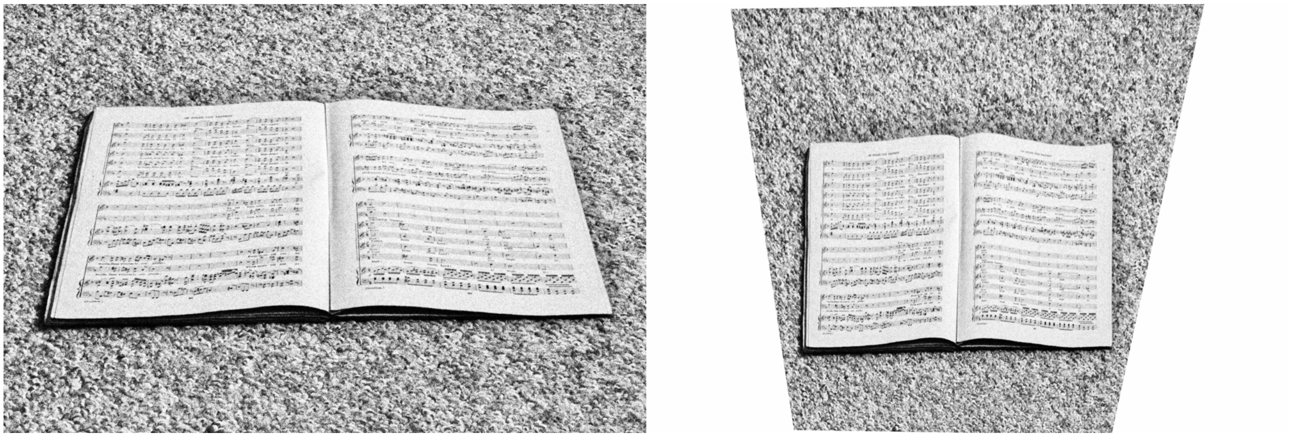


Figure 3: A contrived example for demonstration (original on left, transformed on right): Note particularly how the page is straightened but none of the image edges are parallel or perpendicular.

5 ‘Good Enough’ OMR-enhanced Transcription

Regularising the staff images made OMR approaches to annotation much more viable. To this end, *CANDR* uses a standard pixelwise classification using a convolutional neural network (CNN) and a sliding context window [3]. However, where *CANDR* differs from other approaches is in the purposefully limited scope of its OMR methodology. Often, approaches using neural network methodologies aim for pixel-perfect accuracy for ground truth at a variety of sizes over numerous manuscripts as one small step towards creating accurate reader- or researcher-friendly transcriptions.

Conversely, *CANDR*’s primary input method is manual annotation. As such, creation of OMR-enhanced transcription was desired simply to speed up manual annotation by automating repetitive actions (such as note and staff line positions) which can then be corrected manually. As a result, it is necessary to acknowledge the limitations of both the data and the methodology. Pixel-perfect accuracy requires a pixel-perfect input which is time consuming to create and maintain even with modern pixel classification systems [10]. Such accuracy is not feasible nor necessary with *CANDR*, which aims more narrowly for ‘good enough’ automatic annotation, and a transcriber-centric approach.

CANDR models staff items instead as simplified geometric shapes: clefs, accidentals and syllables are modelled as rectangles, notes are modelled as single points, and staff lines and *divisiones* are modelled as lines. This approach is similar to the methodology of using convex hulls as ground truth, but with the further simplification of those convex hulls constrained to simple 1D and 2D polygons [6]. These items are traced directly onto the straightened staves using the visual editor both as annotation and ground truth for OMR. Then for each stave, average symbol masks are generated for staff lines and other items. *CANDR* conceives of staff lines as a separate layer, onto which other items are placed. Two almost identical CNNs were created: one to classify pixels as staff lines, and another to classify pixels as other items. The separation of these items into two models allows pixels to have two simultaneous and independent classifications: a pixel can exist as a staff line and staff item, and each model can be tuned to better extract its features.

The software that generates these average symbol masks contains parameters for altering its output, such as staff line/*divisione* width and note radius, which were found by hyperparameter tuning but also differ depending on the window size of the classifier. These symbol masks are not a simple binarisation, but the edges are ‘feathered’ to better reflect the average matching of mask to item using supersampling.

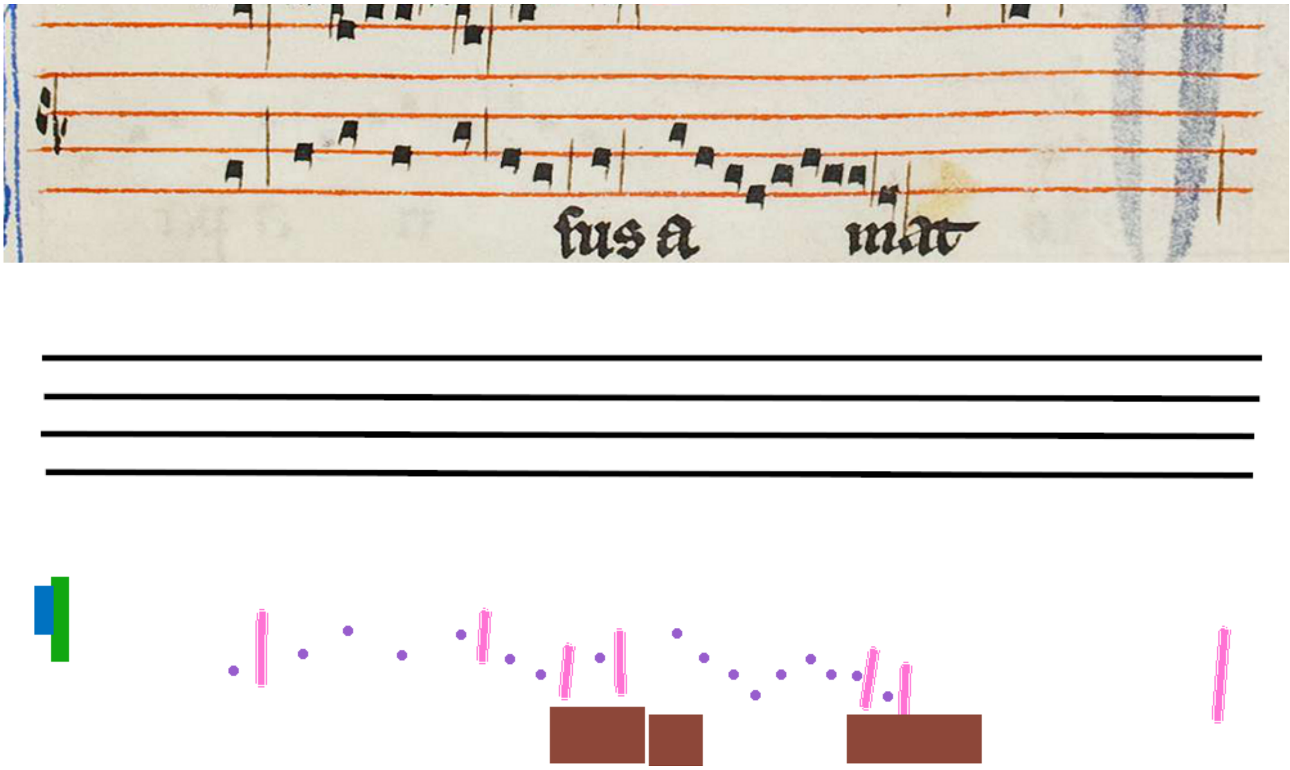


Figure 4: Transformed staff (top), ground truth staff lines mask (middle) and item mask (bottom). The sizes of the mask items are balanced as hyperparameters.

The CNNs were implemented as binary and categorical classifiers for staff lines and staff items respectively. However, rather than binarising the output, the confidence of the prediction for each pixel was multiplied over a radial blur with shape equal to the size of the classifier window, followed by a round of despeckling and Otsu thresholding [8]. The dimensions of most 2D items were detected by calculating the bounding boxes of a connected component analysis, but the 1D staff lines and *divisiones* were instead defined as the coordinates of the furthest extents of each component. To reduce false positives, two further constraints were added: Staff lines must have an angle approaching the horizontal, and must have a width larger than half the width of the image, and, secondly, *divisiones* must approach vertical and must exceed a minimum length.

Notes, however, could not be constructed in the same way as they are modelled as single points, and are often connected into ligatures. Ligatures often resulted in single components on the heatmap. After thresholding, an iterative process of binary erosion was applied to the note heatmap, with the aim of creating as many connected components as possible that were larger than the despeckle parameter. This erosion parameter was found using a binary search.

By not aiming for pixel-perfect accuracy, timing tests indicate that *CANDR*'s use of average symbol masks decreases the time taken for annotation of staves of polyphony threefold over purely manual effort. The fundamental outcome is that, rather than having to annotate staves by tracing items from scratch, transcribing staves of ND polyphony using *CANDR* is now, much more simply and quickly, a process of checking and correcting OMR.

CANDR is released under AGPLv3 at: https://gitlab.com/candr1.

Works Cited

[1] Busse Berger, Anna Maria. "Mnemotechnics and Notre Dame Polyphony" *The Journal of Musicology* 14, no. 3 (1996), 263–298, https://doi.org/10.2307/764059.

[2] Busse Berger, Anna Maria. "Prologue: The First Great Dead White Male Composer" in *Medieval Music and the Art of Memory*. Berkeley: University of California Press, 2005, 9–44, https://doi.org/10.1525/9780520930643-007.

[3] Calvo-Zaragoza, Jorge, Francisco J. Castellanos, Gabriel Vigliensoni, and Ichiro Fujinaga. "Deep Neural Networks for Document Processing of Music Score Images" *Applied Sciences* 8, no. 5 (2018), 654–675, https://doi.org/10.3390/app8050654.

[4] Everist, Mark. "From Paris to St Andrews: the Origins of W1" *Journal of the American Musicological Society* 43, no. 1 (1990), 1–42, https://doi.org/10.2307/831405.

[5] Fujinaga, Ichiro. "Staff Detection and Removal" in *Visual Perception of Music Notation: On-Line and Off-Line Recognition*, ed. Susan E. George. London: IRM Press, 2004, 1–39.

[6] Hajič Jr., Jan, Matthias Dorfer, Gerhard Widmer, and Pavel Pecina. "Towards Full-Pipeline Handwritten OMR with Musical Symbol Detection by U-Nets" in *Proceedings of the 19th International Society for Music Information Retrieval Conference* (ISMIR 2018), 225–232, https://archives.ismir.net/ismir2018/paper/000175.pdf.

[7] Ludwig, Friedrich. *Repertorium Organum Recentioris et Motetorum Vetustissimi Stili* [Halle, 1910], ed. Luther A. Dittmer. New York: Institute of Mediaeval Music, 1964.

[8] Otsu, Nobuyuki. "A Threshold Selection Method from Gray-Level Histograms" *IEEE Transactions on Systems, Man, and Cybernetics* 9, no. 1 (1979), 62–66, https://doi.org/10.1109/TSMC.1979.4310076.

[9] Roesner, Edward. "Notre Dame School" in *The New Grove Dictionary of Music and Musicians*, ed. Stanley Sadie (2nd ed., vol. 18). London: Macmillan, 2001, 202–203.

[10] Saleh, Zeyad, Ké Zhang, Jorge Calvo-Zaragoza, Gabriel Vigliensoni, and Ichiro Fujinaga. "Pixel.js: Web-based Pixel Classification Correction Platform for Ground Truth Creation" in *Proceedings of the 14th IAPR International Conference on Document Analysis and Recognition* (ICDAR 2017), 39–40, https://doi.org/10.1109/ICDAR.2017.267.

[11] Shatri, Elona, and György Fazekas. "Optical Music Recognition: State of the Art and Major Challenges" in *Proceedings of the International Conference on Technologies for Music Notation and Representation* (TENOR 2020/21), 175–184, https://www.tenor-conference.org/proceedings/2020/23_Shatri_tenor20.pdf.

[12] Tardón, Lorenzo J., Isabel Barbancho, Ana M. Barbancho, and Ichiro Fujinaga. "Automatic Staff Reconstruction within SIMSSA Project" *Applied Sciences* 10, no. 7 (2020), 2468–2483, https://doi.org/10.3390/app10072468.

[13] Treitler, Leo. "The Vatican Organum Treatise and the Organum of Notre Dame of Paris: Perspectives on the Development of a Literate Music culture in Europe" in *With Voice and Pen: Coming to know Medieval Song and How it was Made*. Oxford: Oxford University Press, 2003, 68–83, https://doi.org/10.1093/acprof:oso/9780199214761.003.0003.

Annotation of Medieval Music Facsimiles Using “Good Enough” OMR

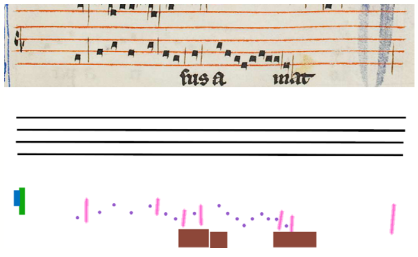
Joshua Stutter
University of Glasgow

Abstract

The *Clausula Archive of the Notre Dame Repertory* (CANDR) is an in–progress PhD project with the aim of cataloguing, transcribing and analysing digital facsimiles of the thirteenth–century repertory commonly termed *Notre Dame polyphony*, and a secondary aim of providing new datasets and analytical tools for studying medieval polyphony. This poster highlights the use in the project of (a) a new methodology for de-skewing facsimile images, and (b) average symbol masks in an OMR–enhanced workflow with an emphasis on creating an OMR workflow that is “good enough” to accelerate the annotation of an image dataset of particularly transitional notation.

Context

- *Notre Dame* (ND) *polyphony* is unfairly elevated & portrayed as the “crucible” of Western art music.
- The field currently relies heavily on the outdated ideas of Friedrich Ludwig’s *Repertorium* (1910).[1]
- CANDR aims to reconstruct ND concordances from scratch
- Usual affine de-skewing methods are unsatisfactory.
- 943 folios of music must be annotated by hand.

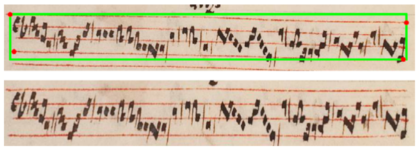


Transformed staff (top), ground truth staff lines mask (middle) and item mask (bottom). The sizes of the mask items are balanced as hyperparameters.

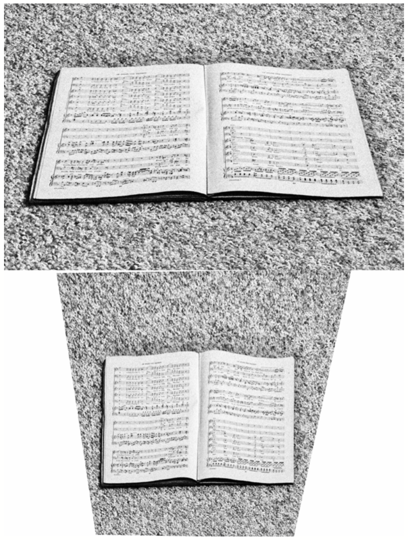
1. Regularisation

CANDR’s regularisation process was designed to remove not only stave shears and rotations but also perspective effects caused by parchment bend and/or offset camera placement. Previous work either: (a) attempts to de-skew staves using shearing and other affine transforms on the image, or (b) considers skewed elements such as staff lines to be a property of the image, electing to model them as smoothed and reconstructed curves.[2, 3] CANDR takes a new approach, using a perspective transform in multiple rounds to de-skew images to yield straightened staves.

First the bounding quadrilateral of a system is manually annotated on a facsimile image and an inverse perspective transform matrix is calculated that can transform that quadrilateral into a straightened rectangle. A second bounding quadrilateral for each stave on that system can then be manually annotated upon the straightened system and the same process applied to create individual straightened and cropped stave images. This process is integrated into the annotation workflow of CANDR and is calculated on–the–fly: annotating a system or stave creates a new object in the database along with its bounding quadrilateral. Viewing a system or stave in the visual editor is to view that item transformed into a straight rectangle using this transform. This method (a) mitigates the effects of parchment bend significantly and (b) can eliminate not only rotation and skew but also perspective effects in the facsimile. There is also the added advantage of presenting to the user a clear and straightened image.



A real–world example: Before transform of four–line staff with staff boundaries and target rectangle marked (top), and after transform (bottom).



A contrived example for demonstration (original on top, transformed on bottom): Note particularly how the page is straightened but none of the image edges are parallel or perpendicular.

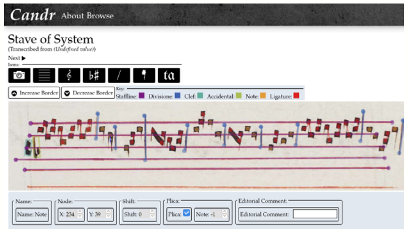
2. “Good enough” OMR

CANDR uses a standard pixelwise classification using a convolutional neural network (CNN) and a sliding context window.[4] However, the primary input method of the 943 folios is manual annotation and as such, creation of OMR–enhanced transcription was desired simply to speed up manual annotation by automating repetitive actions which can then be corrected manually. Rather than aiming for pixel–perfect accuracy,[5] CANDR models staff items as simplified geometric shapes: clefs, accidentals and syllables are modelled as rectangles, notes are modelled as single points, and staff lines and *divisiones* are modelled as lines. This approach is similar to the methodology of using convex hulls as ground truth, but with the further simplification of those hulls constrained to simple 1D and 2D polygons.[6] For each stave, average symbol masks are generated for staff items.

The software that generates these masks contains parameters for altering its output, such as line width and note radius, which were found by hyperparameter tuning. These symbol masks are not a simple binarisation, but the edges are “feathered” to better reflect the average matching of mask to item using supersampling. The dimensions of most 2D items were detected by calculating the bounding boxes of a connected component analysis, but the 1D items were instead defined as the coordinates of the furthest extents of each component.

Notes however, could not be constructed in the same way as they are modelled as single points, and are often connected into ligatures. Ligatures often resulted in single components for multiple notes. To ameliorate this issue, after thresholding an iterative process of binary erosion was applied to the analysis with the aim of creating as many connected components as possible that were larger than the despeckle parameter. This was found using a binary search.

Timing tests indicate that CANDR’s use of average symbol masks decreases the time taken for annotation of staves of polyphony threefold over purely manual effort. The fundamental outcome is that, rather than having to annotate staves by tracing items from scratch, transcribing staves of ND polyphony using CANDR is now a process of checking and correcting OMR.



Stave of System
(Transcribed from (*Undefined value*))

Key: Staffline: 🟪 Divisione: 🟦 Clef: 🟩 Accidental: 🟩 Note: 🟧 Ligature: 🟥

Name: Name: Note
Node: X: 234 Y: 39
Shift: Shift: 0
Plica: Plica: ☑ Note: -1
Editorial Comment: Editorial Comment:

A fully–annotated stave: horizontal stafflines (purple), C clef (teal), vertical “divisiones” (blue), notes and ligatures (orange and red respectively).

[1] Ludwig, Friedrich. *Repertorium Organum Recentioris et Motetorum Vetustissimi Stili*. Ed. Dittmer, Luther. New York: Institute of Mediaeval Music, 1910

[2] Fujinaga, Ichiro. “Staff Detection and Removal.” *Visual Perception of Music Notation: On-Line and Off-Line Recognition*. Ed. George, Susan E. London: IRM Press, 2004. 1-39.

[3] Tardón, Lorenzo J., Isabel Barbancho, Ana M. Barbancho, Ichiro Fujinaga. “Automatic Staff Reconstruction within SIMSSA Project.” *Applied Sciences* 10 7 (2020): 2468-2483.

[4] Calvo-Zaragoza, Jorge, Franciso J. Castellanos, Gabriel Vigliensoni, Ichiro Fujinaga. “Deep Neural Networks for Document Processing of Music Score Images.” *Applied Sciences* 8 5 (2018): 654-675.

[5] Saleh, Zeyad, Ké Zhang, Jorge Calvo-Zaragoza, Gabriel Vigliensoni, Ichiro Fujinaga. “Pixel.js: Web–based Pixel Classification Correction Platform for Ground Truth Creation.” *Proceedings of the 14th IAPR Internation Conference on Document Analysis and Recognition*. Kyoto: Springer LNCS, 2017. 39-40.

[6] Hajič Jr., Jan, Matthias Dorfer, Gerhard Widmer, Pavel Pecina. “Towards Full–pipeline Handwritten OMR with Musical Symbol Detection by U-Nets.” *Proceedings of the 19th ISMIR Conference*. Paris, 2018. 225-232.