# Musical Stylometry, Machine Learning, and Attribution Studies: A Semi-Supervised Approach to the Works of Josquin

Andrew Brinkman[*1] Daniel Shanahan, [*2] Craig Sapp[#3]

[*]*School of Music, Louisiana State University, Baton Rouge, Louisiana, USA*
[#]*CCARH/PHI, Stanford University*
[1]abrink4@lsu.edu, [2]dshanahan@lsu.edu, [3]craig@ccrma.stanford.edu

## ABSTRACT

Compositional authorship is often assigned though factors external to the musical text, such as biographical records and surveys of source attributions; however, such methodologies often fall short and are potentially unreliable. On the other hand, determining compositional authorship through internal factors—through stylistic traits of composers derived from the music itself—is often fraught with errors and biases. One of the underlying assumptions in the field of stylometry is that, while it is difficult for humans to perform a truly unbiased analysis of authorial attribution, computational methods can provide clearer and more objective guidelines than would otherwise be apparent to readers or listeners, and thus might provide corroboration or clues for further investigation. This paper discusses machine-learning approaches for evaluating attribution for compositions by Josquin des Prez. We explore musical characteristics such as melodic sequences, counterpoint motion, rhythmic variability, and other entry measures to search for features inherent to a composer's works or style, and we hope that employing such an approach—one that explicitly states which factors led to the decision-making process—can serve to inform scholars looking at other works and composers.

## I. INTRODUCTION

Perhaps more than any other composer, Josquin has generated the most scholarly debate regarding the attribution of his works: those that were at one point seen as firmly rooted in the composer's *oeuvre* are now frequently discredited and are being reconsidered. This controversy stems from a number of factors, including the lack of primary source material. In the case of the *New Josquin Edition*, a collection of Josquin's music sorted by various factors (including controversial attribution levels), it often designates a work's authorship by the amount of times that work appears within primary collections (Wegman, 2008).

Only a few other factors contribute to determining attribution within these works. These primarily include whether or not a work appears in direct reference to the composer in biographical accounts of the time. Additionally, of the over 400 works attributed to the composer across multiple genres, more than 200 of them are considered to have questionable levels of attribution (Rodin and Sapp, 2016). This, coupled with uncertainties about sources for printed editions of his works (particularly for posthumous publications), increases the likelihood of incorrect attribution. Outside of the appearance of works within primary sources, there is no consistently reliable process for determining attribution. While many musicologists have suggested a variety of methods to help determine authorship without the use of primary sources, these methods are unreliable and frequently contested (see Sparks 1971; Wegman, 2008, and Jas, 2014). The Josquin Research Project (JRP) has the most rigorous and systematic in its approach to identifying possibly problematic attributions, doing so based solely on early manuscript sources. In the JRP, an attribution level of "1" means the piece is attributed to the composer in at least two early manuscripts, whereas a piece with an attribution level of "2" is referred to in only one early source. Levels 3–6 represent decreasing levels of certainty about pieces for which there are no early manuscript sources.[1]

Stylistic features are debated between scholars as the composer's style may have likely changed over the course of their career, and what were once thought to be idiosyncrasies are often shown to appear more broadly in the repertoire. This study tries to mitigate this issue through multiple machine-learning models. By utilizing both high-level features suggested by these musicologists as well as several low-level features (such as note-to-note transitions and correlations to securely-attributed works), we hope to be able to provide more likely starting points for researchers, as well as a model that will ideally grow into a tool for more general attribution studies in music research.

## II. MUSICAL STYLOMETRY

Stylometry has been used in the humanities to better determine and understand the authorship of works ranging from the *Federalist Papers* (Motseller & Wallace, 1964) to the works of Shakespeare (Merriam & Matthews, 1994), and more recently the pseudonymous work of J.K. Rowling (Juola, 2015). Similar statistical approaches have even been used to analyze what might be coerced confessions from suspects (and possibly forged confessions by law enforcement) in criminal trials (Kenny, 2013). A central principle of computational stylometry is that the low-level features that are often undetectable to even the most experienced reader can serve as a sort of "authorial fingerprint", distinguishing an author from their contemporaries and colleagues (see Love, 2002).

This method is viewed—often justly—with a certain level of skepticism. Early approaches had many results that were eventually disproven, such as A.Q. Morton's analysis of the works of Joyce (see Levitt, 2001). Recent models have improved with advances in statistical methods and machine-learning techniques. Models tend to use a mixture of high-level features, such as word preferences (e.g. the universal "he", "she", "he/she", or the singular "they") and intentional coinages, such as those employed by Coleridge (see Love, 2002; pp.107–108), as well as low-level features, such as the

[1] for more information, see http://josquin.stanford.edu/about/attribution.

transitional probabilities of certain words, and the correlation of the word frequencies to a corpus of confirmed authorship. Music researchers have occasionally used such techniques to determine compositional authorship. For example, Backer and Kranenburg (2005) used a linear discriminant transformation to examine the features of works by J.S. Bach and J.L. Krebs (see also Kranenburg, 2007), and Bellman (2011) employs computational methods to quantify musical style engaging with stylometric questions throughout his study. Speiser and Gupta (2013) use principal components analysis to train classifiers (similarly to how this paper approaches the problem). The current study makes use of these previously implemented approaches to better understand the authorship of the works of Josquin.

### A. High-Level Musical Features

This study makes use of what we might consider "high-level" features, meant to accompany the currently employed "low-level" features. The former category includes the presence of 9-8 suspensions, as well as the rate of occurrence of the specific types of contrapuntal motion: parallel, similar, contrary, and oblique. Many other commonly-discussed high-level features are not employed in the current study. For example, musicologists often debate the uniqueness of the texture around cadence points in Josquin's music, arguing that textures become sparser as cadences approach (Wegman, 2008). Such features can be difficult to quantify, and therefore difficult to operationalize. Other features commonly discussed but left out of the current study include "relaxed" contrapuntal writing and an overreliance on repetition of material. These both require a level of interpretation that we decide not to employ for the current study.

Additionally, a feature commonly referred to as the *Satzfehler* has been discussed and debated. Originally coined by Helmuth Osthoff (Osthoff, 1962), the *Satzfehler* was considered to be a possible error in voice-leading made by Josquin during his early years, in which leading-tone motion occurred between two voices simultaneously and involved one voice dipping down from the tonic to the leading-tone and then ascending back to the tonic while another voice preemptively performs the leading-tone before the first voice has had a chance to resolve. Later Edgar Sparks argued that this device was not an error but was simply a common feature used, not only by Josquin, but also by many of his contemporaries (Sparks, 1976). While the appropriateness of this technique is still being questioned, it is one of the few techniques that have explicitly been applied to Josquin's music as an identifier of his compositional style, and have been documented at length (Wegman, 2008). For the current study, however, we decided to exclude it from our searching, as once again there were many flexibilities to the definition, and computational searching yields too many false positives.

### B. Low-Level Musical Features

For low-level features, we examined note-to-note transitions by scale degree. Admittedly, the notion of scale degree does not easily map onto the music of the Renaissance, but it nevertheless provides a concise method for looking at relationships between notes. These were then examined along with other similar low-level features including rhythmic variability, measured with the normalized pairwise variability index (see Grabe and Low, 2002; Daniele and Patel, 2004; Patel and Daniele, 2013), the average entropy of each melodic voice, and the correlation to the pitch distributions in secure Josquin works.

**Table 1. Features contributing to the models.**

| High-Level | Low-Level | Metadata |
|---|---|---|
| 9-8 suspensions | average melodic entropy (per voice) | primary listed composer |
| oblique motion | nPVI | title |
| contrary motion | pitch-range correlation to secure Josquin works | JRP attribution level |
| similar motion | note-to-note transition probabilities (by scale degree) | genre |
| parallel motion | | |

## III. BETWEEN-COMPOSER COMPARISON

We compare the music of Josquin to that of other composers in a range of styles. This helps to identify features that represent differences between styles, are common to a particular style, or are unique to a particular composer. A comparison between Josquin and J.S. Bach's four-part chorales is used as an initial calibration comparison. Most any untrained listener should be able to distinguish between these two repertories separated by two centuries. Slightly more difficult are comparisons to Ockeghem and Du Fay, who predate Josquin by one and two generations. These comparisons are more difficult for the average listener, but relatively easy for experienced listeners of Renaissance music. Finally, challenging comparisons are made to de Orto, and La Rue. These composers are Josquin contemporaries who all compose in a common style, and their works are often misattributed amongst each other.

### A. Dimensionality Reduction

Before training a classifier, we employed principal component analysis (PCA) to reduce the 53 input features into a more manageable number of dimensions. This maintains as much variance between features as possible and is used to drop dimensions containing little information. High-level feature comparisons seem to be more salient between genres and epochs, whereas low-level features function best within them. To anticipate the results of this portion of the study, comparisons between disparate styles (Josquin and Bach) were actually hindered by the inclusion of low-level features, and the classifiers performed better when only provided the high-level information. In contrast, low-level features do increase the separation abilities of PCA and can train more accurate classifiers when provided with whose music in style similar to that of Josquin.

## B. Josquin versus Bach

Before training and testing a model on composers with similar styles to Josquin, we examine differences between the works of Josquin and Bach chorales. If musical features were to have difficulties between two composers, then it would be logical to assume that progression to more similar style comparisons would prove to be problematic, let alone be suitable for authorship studies. High-level features were used to compare the works of Josquin (those listed as secure according to JRP attribution assignments) and the Bach chorales available in kern format.[2] When applied to only the high-level features, PCA shows that most of the variance can be accounted for within the first two principal components. As can be seen in Figure 1, the elbow might argue that the model would be feasible with only these two components. We instead opted to include the first five components, which account for 85% of the variance.

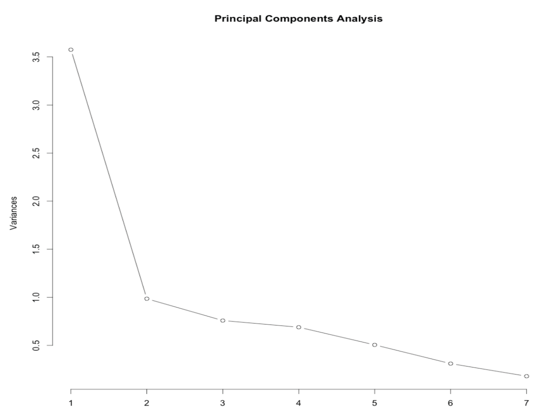


**Figure 1. Variance accounted for in each principal component when comparing Bach and Josquin.**

Figure 2 illustrates the strength of each high-level feature in this model. The largest amount of variance seems to be accounted for by the contrapuntal and melodic techniques (9-8 suspensions, similar, parallel, and oblique motion), as well as rhythmic variability (the nPVI of the entire piece, notated as nPVI_entire). This will makes sense to anyone who has spent any amount of time listening to these two composers.

A relatively clear separation can be seen in Figure 3, with the model able to discern the music of Bach from Josquin's music quite easily and from relatively few features. Oblique motion, pitch entry and pitch correlations are the best base features to distinguish between these two repertories. As we will see with our classifiers, these principal components can be used to train a model that can easily discern Bach from others composers being evaluated (Table 2).

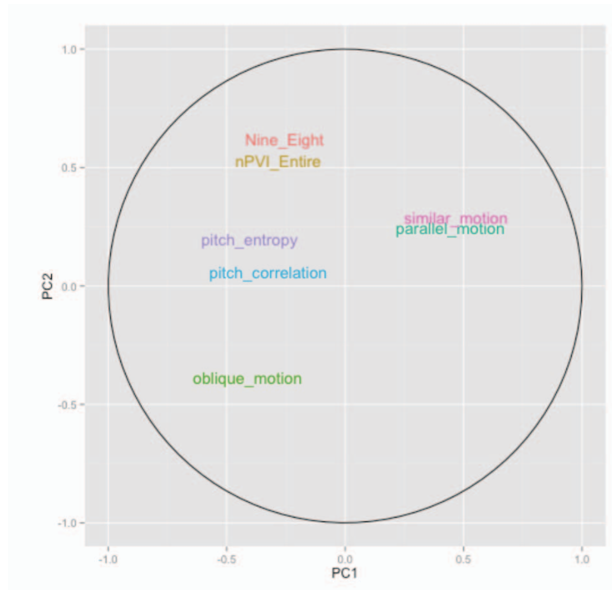


**Figure 2. PC coefficients associated with the variables between the Josquin and Bach datasets.**


**Figure 3. The separation of Bach and Josquin given only the two PCs that account for the most amount of variance (full model uses 5 PCs).**

## C. Josquin and Du Fay

A more difficult task involves comparisons to works by composers closer to Josquin's lifetime (*c*1450–1521). Guillaume Du Fay (1397–1474) provides a good starting point, as the two are separated by two generations. Low-level features were included for this comparison to facilitate separation. Figure 4 shows the contribution of features to the first two principal components. Repetition of the fifth scale degree seems to account for the highest level of variance between these two composers as seen by the "t5_5" label in the figure with a correlation around 0.25 for the first principal components axis.

[2] https://github.com/craigsapp/bach-371-chorales. JRP scores can most conveniently also be downloaded from Github at https://github.com/josquin-research-project/jrp-scores.

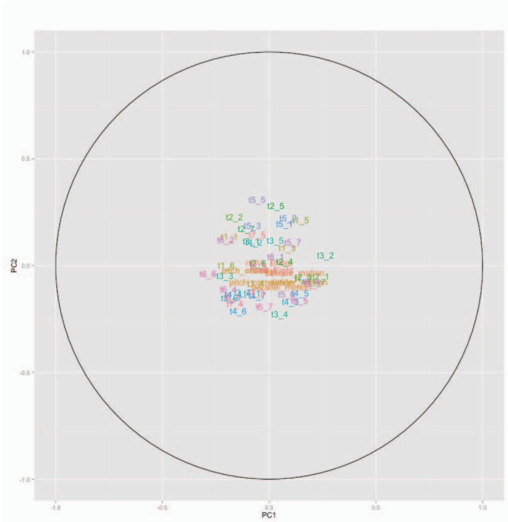


**Figure 4. PC coefficients associated with the variables between the Josquin and Du Fay datasets.**

Looking at the biplot in Figure 5 for the separation, we see that the first two PCs are not able to separate the two composers quite as well as with Josquin/Bach. The model relies on more features (the features in addition to the high-level), so each PC accounts for less variance.

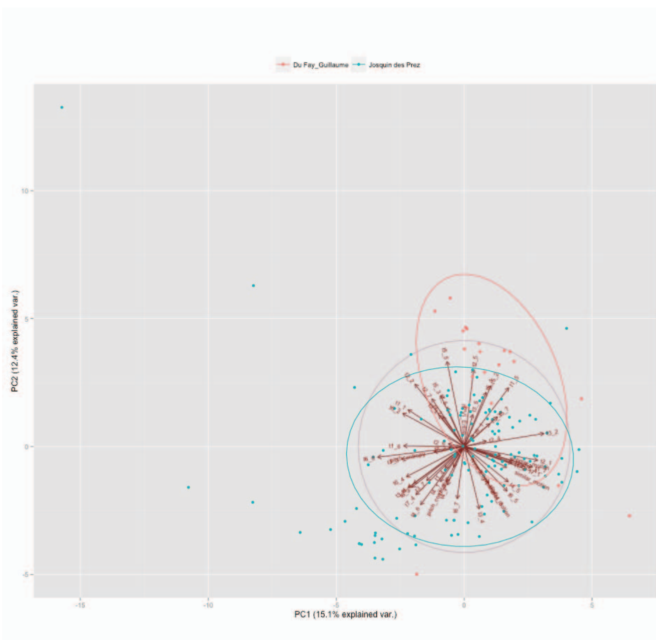


**Figure 5. The separation of Josquin and Du Fay visualized using only the first two PCs.**

## D. Josquin and Ockeghem

Johannes Ockeghem's (*c*1425–1497) style is yet more similar to Josquin's than Du Fay or Bach. Yet expert listeners can distinguish relatively easily between the styles of Josquin and Ockeghem. As with Du Fay, repetition of the fifth stands is a distinguishing feature between the two composers, and (unlike Du Fay) the treatment of the second scale degree accounts for a decent amount of the variance in the top two principal components.

We might predict that our classifiers would score fairly well in classifying between these two composers, although less well than Bach, and perhaps less clearly than Du Fay.

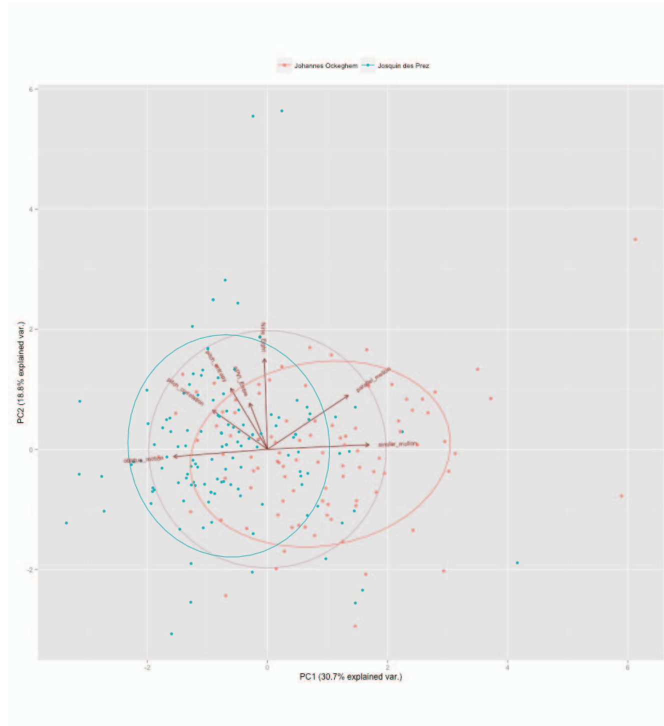


**Figure 6. The separation of Josquin and Ockeghem visualized using only the first two PCs.**

## E. Josquin and de Orto

Mabrianus de Orto (1469–1529) is more closely a contemporary of Josquin's, and their styles even closer than the previously mentioned composers. Interestingly, the repetition of the fifth continues to play a strong role, but with a negative correlation. Melodic sequences involving the fourth (3 → 4 and 4 → 6) strongly correlate between the two composers. It is likely that the model would perform fairly poorly on de Orto, and this might be hinted at in Figure 7's biplot. In this case the first PCs describing de Orto's works is nearly completely encompassed within that of Josquin's works.

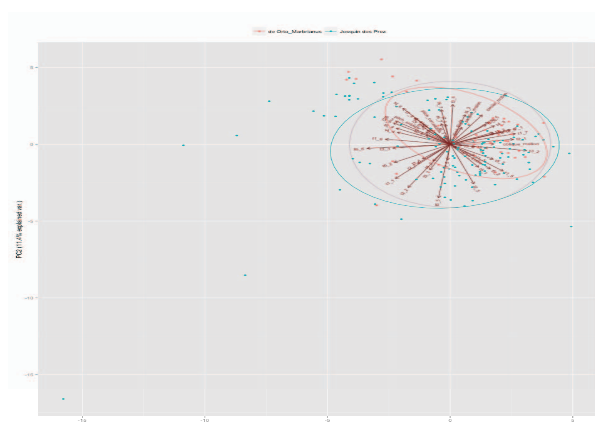


**Figure 7. The separation of Josquin and de Orto visualized using only the first two PCs.**

### F. Josquin and La Rue

Pierre de la Rue (c1452–1518) is more problematic than de Orto. His works are most commonly mistaken for Josquin's (and vice versa), and the two are the most stylistically similar of all of the composer's examined in the present study. Figure 8, shows an interesting pattern: the repetition of the fifth is anti-correlated, and the strongest correlation coefficients are provided by 3 → 4 scale degree motions.

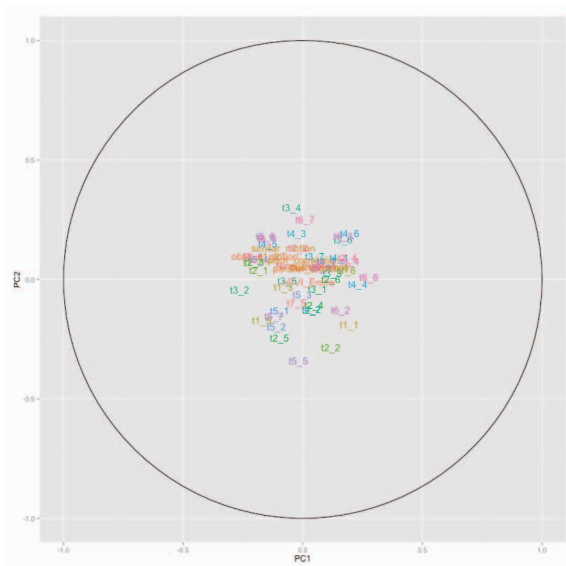


**Figure 8. First two PC coefficients for Josquin and La Rue features.**

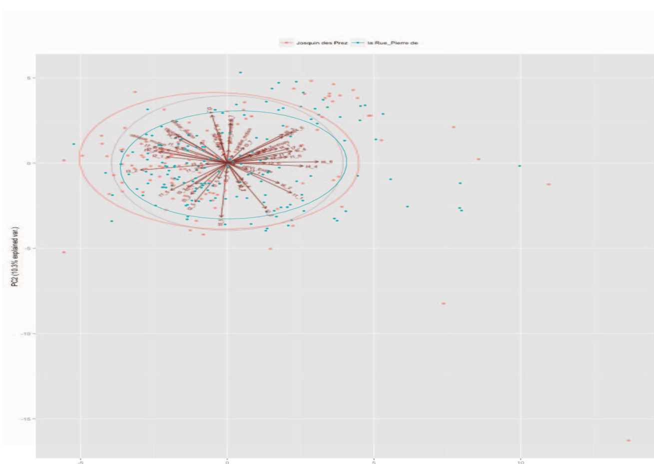

**Figure 9. The separation of Josquin and La Rue visualized using only the first two PCs.**

Figure 9 shows the biplot for Josquin and La Rue. The two composers complete overlap each other (and thus the first two PCs cannot be used alone to distinguish between the composers at all). This figure illustrates why many researchers have mistakenly cross-attributed the works of both composers.

## IV. CLASSIFIER PERFORMANCE

While the previous section focused on binary comparisons between Josquin and another composer, attribution does not exist in such tidy laboratory conditions. Ideally, all composers should be fed into a classifier. In order to achieve this goal, the current study uses the results of the principal component analysis as a way of training a multi-composer classifier. As this study is primarily exploratory in nature, we opted to train three types of classifiers, and compare the results. A future goal, as will be discussed below, would be to employ an ensemble-learning model, in which multiple models are used and a decision tree is implemented to take the results of the model that best performs in a given context. The current study employs a k-nearest neighbor (KNN), a support vector machine, and a decision tree (as a way of better understanding the role of each feature).

### A. K-Nearest Neighbor Classifier

A principal component analysis was performed on all of the above composers at once, using 48 features. In order to account for 85% of the variance (the standard used above), 27 PCs were needed to create the model. A two-dimensional representation of the separation between the first two PCs can be seen below in Figure 10. Notice that the first two PCs can be used to anticipate the similarity of style between the composers. Bach (red) is well separated from all of the Renaissance composers, while Du Fay (green) is the second more differentiated. The first two PCs cannot differentiate by themselves between the other composers' music (Ockeghem, de Orto, La Rue and Josquin).

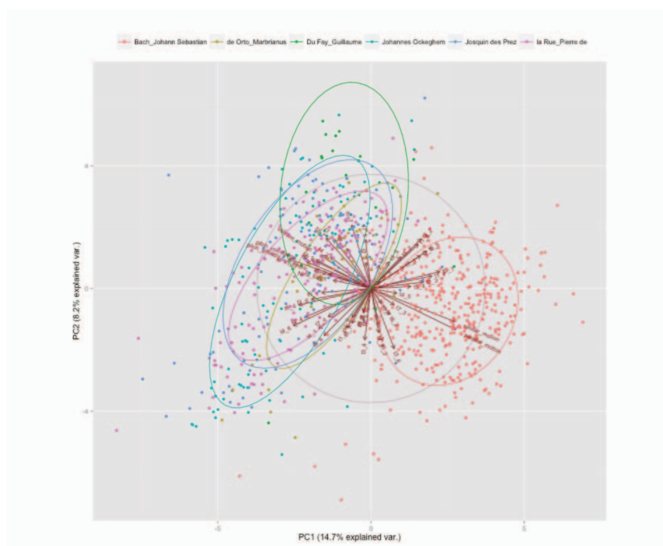


**Figure 10. The separation of all composers visualized using only the first two PCs (this model incorporated 27 principal components).**

An 80/20 split was used (the model was trained on 80% of the data and tested on 20%), with $k$=6. A confusion matrix is provided below in Table 2. Overall, the model performed quite well in discerning Bach from the other composers, indicating once again that very little is needed in order to determine pieces by genre and time-period. One Bach chorale was mistakenly assigned to de Orto, and another to Josquin. The model correctly predicted de Orto 38.9% of the time, but assigned a couple of pieces each to Ockeghem and Josquin. The model's identification of de Orto's works was most confused by La Rue, mistakenly assigning de Orto pieces to him a third of the time.

Du Fay was most confused with Ockeghem, which makes sense, given their relative proximity in time period and similar compositional styles. Interestingly, however, this was not a bidirectional confusion: Ockeghem pieces were rarely confused for those written by Du Fay, and instead were more likely to be ascribed to La Rue (20% of the time). The model did surprisingly well with discerning Josquin from La Rue, which was predicted to be the most difficult comparison. Although 21.2% of Josquin pieces were confused for La Rue pieces, the model guessed correctly on 60.6% of the pieces. For La Rue, the model was accurate 80.6% of the time, and only ascribed pieces to Josquin 16.7% of the time. While, at first sight some of these accuracy levels might seem to be a little low—we had hoped for at least a 75% accuracy with Josquin—it should be noted that each composer performed significantly better than the baseline (which would be around 16.7%, given a uniform random comparison between six composers).

**Table 2. Confusion matrix for KNN model. Results are in a percentage of responses. The correct answer is listed in the first column, and predictions are listed in the top row.**

| | Bach | de Orto | Du Fay | Ock. | Josquin | La Rue |
|---|---|---|---|---|---|---|
| **Bach** | 94.5 | .9 | 0 | 0 | .9 | 3.6 |
| **de Orto** | 5.6 | 38.9 | 0 | 11.1 | 11.1 | 33.3 |
| **Du Fay** | 14.3 | 0 | 42.9 | 28.6 | 0 | 14.3 |
| **Ockeghem** | 0 | 6.7 | 3.3 | 70 | 0 | 20 |
| **Josquin** | 0 | 9 | 3 | 6.1 | 60.6 | 21.2 |
| **La Rue** | 0 | 2.8 | 0 | 0 | 16.7 | 80.6 |

### B. Support Vector Machine

We next trained a support vector machine (SVM) classifier, intended to improve on the results of the first. KNN have nice properties such as automatic adaption to non-linear data, but it has to be tuned carefully and is sensitive to bad features. SVMs are useful for segmenting data in high-dimensional spaces. The SVM used a radial basis function kernel which allows for a classification boundaries that might curved, rather than linear. The same PCs were used as in the k-nearest neighbor classifier.

The results, which can be seen in Table 3, are limited in scope: there was a limited number of pieces in the random test set available for composers such as de Orto and Du Fay. Nevertheless, the results were consistent across multiple tests. Interestingly, the model does not seem to do much better when predicting Josquin, but the accuracy improves slightly with La Rue, while the Bach prediction remains quite high. Surprisingly, the SVM performs similarly to the KNN method, with ratings close to the assignments in Table 2.

**Table 3. Confusion matrix for SVM using a radial basis function. Results are in percentage of responses. The correct answer is listed in first column, top row lists predicted composer.**

| | Bach | de Orto | Du Fay | Ock. | Josquin | La Rue |
|---|---|---|---|---|---|---|
| **Bach** | 98.5 | 0 | 0 | 0 | 0 | 1.5 |
| **de Orto** | 33.3 | 33.3 | 0 | 0 | 0 | 33.3 |
| **Du Fay** | 50 | 0 | 25 | 25 | 0 | 0 |
| **Ockeghem** | 6.7 | 0 | 0 | 60 | 13.3 | 20 |
| **Josquin** | 16 | 0 | 4 | 4 | 60 | 16 |
| **La Rue** | 0 | 0 | 0 | 0 | 12.9 | 87.1 |

### C. Future Work

Future work will examine the specifics of each feature more thoroughly as a way of identifying the differences between composers. For example, the features of all of the composers studied were placed within a decision tree model to better understand how such aspects might contribute to creating better models in the future. As can be seen below, the most important features in discerning the four composers seem to be the employment of similar motion, as well as treatment of the sixth scale degree, followed by motion of scale degree two to scale degree five. While the decision tree below divides the pieces into eight, rather than six, compositional styles, it allows us to better understand the differences the constitute each composer's style and provide some context for the above PCA and subsequent classifiers.

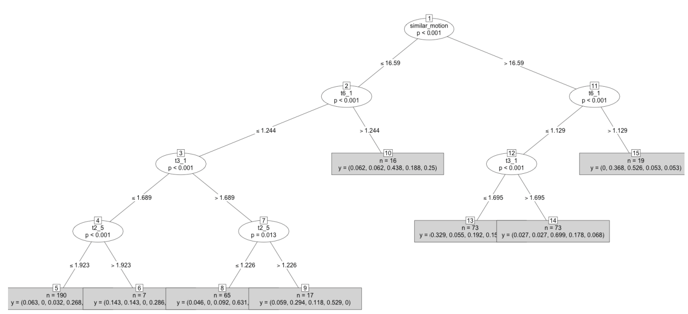

**Figure 11. Decision tree analysis diagram**

## V. CONCLUSION

The current study attempts to demonstrate the applicability of machine-learning techniques to the question of authorship in Josquin's music. There are obviously quite a few points that need to be further discussed. For example, genre matters. Composers wrote differently for masses than they did for motets, yet the current study (primarily to have as much statistical power as possible) treats each composer's output as a uniform stylistic output. Secondly, it assumes that compositional style is undergoes minimal levels of change throughout the course of a composer's career. Future work hopes to be able to use these elements, although the sources are often sparse and apocryphal. Additionally, we hope to examine more features in future work, including harmonic entropy, rhythmic transition probabilities, and the incorporation of the aforementioned *Satzfehler*. Once these are implemented we hope to provide a more accurate model that can serve as a tool for researchers investigating this music.

## ACKNOWLEDGMENTS

We would like to thank Jesse Rodin and the entire team at the Josquin Research Project for all of the work they do on the project, and Blake Howe for his helpful advice and insight into the music of Josquin.

## REFERENCES

Backer, E., & Kranenburg, P. van. (2005). On musical stylometry—a pattern recognition approach. *Pattern Recognition Letters*, 26(3), 299–309.

Bharucha, J. J. (1991). Pitch, harmony, and neural nets: A psychological perspective. In P. Todd & G. Loy (Eds.), *Music and Connectionism* (pp. 84–99). Cambridge, MA: MIT Press.

Daniele, J. R., & Patel, A. D. (2013). An Empirical Study of Historical Patterns in Musical Rhythm: Analysis of German & Italian Classical Music Using the nPVI Equation. *Music Perception*: *An Interdisciplinary Journal*, (31/1). pp. 10–18.

Grabe, E., & Low, E. L. (2002). Durational variability in speech and the Rhythm Class Hypothesis. In Laboratory Phonology, (7). pp. 515–546. Berlin: Mouton de Gruyter.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation*. Cambridge, MA: MIT Press.

Jas, E. (2014). What Other Josquin? *Early Music History*, 33, 109–142.

Juola, P. (2015). The Rowling Case: A Proposed Standard Analytic Protocol for Authorship Questions. *Digital Scholarship in the Humanities*, 30 (suppl 1), pp. i100–i113.

van Kranenburg, P. (2006). Composer attribution by quantifying compositional strategies. In ISMIR Proceedings, pp. 375–376.

Levitt, M. (2001). Joyce, Yet Again. *Journal of Modern Literature*, 24(3), 507–511.

Love, H. (2002). *Attributing Authorship*. Cambridge: Cambridge University Press.

Mosteller, F., & Wallace, D. (1964). *Inference and Disputed Authorship: The Federalist*. Addison-Wesley.

Osthoff, H. (1962). Josquin des Prez, in MGG1, vol. 7. pp. 190–214.

Parncutt, R., & McPherson, G. E. (Eds.) (2002). *The science and psychology of music performance: Creative strategies for teaching and learning*. New York: Oxford University Press.

Patel, A. D., & Daniele, J. R. (2003). An empirical comparison of rhythm in language and music. *Cognition*, (87/1).

Rodin, J, & Sapp, Craig. (2010). The Josquin Research Project. http://josquin.stanford.edu, accessed on 5/5/2016.

Speiser, Jacqueline, & Gupta, Vishesh. "Composer Style Attribution". Project Report for CS 229: Machine Learning, Stanford University; 2013. cs229.stanford.edu/proj2013/CS229-FinalProject.pdf

Sparks, E. H. (1971). Problems of authenticity in Josquin's motets. Proceedings of the International Festival-Conference. edd. E.E. Lowinsky & B.J. Blackburn (London 1976), pp. 345–59.

Wegman, R. C. (2008). The Other Josquin. Tijdschrift van de Koninklijke Vereniging Voor Nederlandse Muziekgeschiedenis, 58(1), 33–68.