A CORPUS-BASED ANALYSIS OF SYNCOPATED PATTERNS IN RAGTIME

Phillip B. Kirlin
Department of Mathematics and Computer Science, Rhodes College
kirlinp@rhodes.edu

ABSTRACT

In this paper, we build on and extend a number of previous studies of rhythmic patterns that occur in ragtime music. All of these studies have used the RAG-C dataset of approximately 11,000 symbolically-encoded ragtime pieces to identify salient rhythmic patterns in the corpus and qualify how they are used. Ragtime music is distinguished from other musical genres by frequent use of syncopation, and previous computational studies have confirmed a number of musicological hypotheses regarding the use of syncopated patterns in ragtime compositions. In this work, we extend these studies to investigate further questions involving the use of syncopation. Specifically, we introduce a new methodological framework for processing the RAG-C dataset and confirm that experiments from previous studies obtain similar results using the new methodology. We investigate the use of the common "short-long-short" syncopated pattern in different time periods and present new results detailing its use by three well-known ragtime composers. We describe how the use of other syncopated patterns has evolved over time and the different distributions of patterns that result from those changes. Lastly, we present novel results identifying statistically significant patterns in the way composers varied the amount of syncopation in consecutive measures in compositions.

1. INTRODUCTION

In this work, we present an analysis of the salient rhythmic patterns that occur in ragtime piano music and quantify how the use of these patterns has changed over time and varies between composers. In particular, we illustrate that the specific sequences of syncopated patterns found in ragtime music and the ways in which they are ordered are not due to chance, but due to deliberate choices made by the composers. We argue that understanding and quantifying the musical choices made by composers is crucial to creating and improving the performance of various music information algorithms, including those for genre classification and algorithmic composition.

Recently, a number of corpus-based studies of ragtime music have been published [1–4] that use a dataset known as the RAG-collection (RAG-C), a corpus of over 11,000 MIDI files first introduced by Volk and De Haas [1]. This dataset presents particular challenges to use due to its heterogeneous nature: it contains compositions from a wide variety of eras and composers, includes numerous ragtime styles (some of which could be argued are not ragtime at all), and has no standard method for encoding the music in MIDI format: some files are derived from live performances, while others are sequenced from sheet music.

Previous studies using the RAG-C corpus have identified the common usage of certain rhythmic patterns in ragtime, but have varied in their techniques in processing the corpus and interpreting the rhythmic patterns located. We present a new methodology for analyzing the corpus while taking care to confirm that our results align with previously-published studies.

Our contributions are as follows. First, we illustrate the feasibility of using automated algorithmic techniques to extract rhythmic patterns from a collection of MIDI files. We also argue for why our particular techniques work given the heterogeneous nature of the RAG-C dataset, especially involving the different time signatures present in the collection. Second, we extend previous corpus-based studies of ragtime music to illustrate the importance of specific kinds of syncopated patterns across the entire corpus, and also in subsets segmented by era and by composer. Since our methodological techniques are slightly different from those used in previous work, we confirm a number of earlier results and then extend them with new experiments and findings. Third, we show that additional patterns emerge when we study the way composers choose to order the amount of syncopation in successive measures: we statistically illustrate that this is done in a particular, deliberate manner. All the code for the experiments described here is publicly available. [1]

2. RAGTIME AND SYNCOPATION

In music, *syncopation* occurs when notes that a listener would expect to occur on strong beats in a measure are shifted to weak beats. Syncopation is particularly identified with *ragtime* music; while various definitions of ragtime exist, the unifying characteristic is the presence of certain varieties of syncopated rhythms [5]. Though in the modern era ragtime is often thought of as a form of music

[1] https://github.com/pkirlin/ragtime-ismir-2020

 Attribution: P. Kirlin, "A Corpus-Based Analysis of Syncopated Patterns in Ragtime", in *Proc. of the 21st Int. Society for Music Information Retrieval Conf.,* Montréal, Canada, 2020.

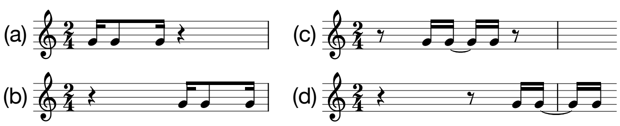


Figure 1. Different versions of the 121 pattern. Variations (a) and (b) are untied; (c) and (d) are tied.

restricted to the piano, during ragtime's heyday of roughly 1890–1920, this style of music was composed for all kinds of instrumental ensembles as well as in song form [6]. After 1920, ragtime fell out of compositional favor, though one can still find many rags composed during the modern era.

As syncopation is the defining characteristic of ragtime, it is natural to study how the use of syncopated patterns has changed over time. Ragtime scholars argue that a number of specific syncopated rhythmic patterns that composers gravitated towards changed in their frequency of use during the original ragtime era, with particularly drastic shifts occurring around the turn of the century. In particular, musicologists often focus on the importance of the "short-long-short" or "121" pattern. This pattern occurs with various note durations, but is usually found in the form 𝅘𝅥𝅮 𝅘𝅥 𝅘𝅥𝅮 in $\frac{4}{4}$ or $\frac{2}{2}$ time signatures and 𝅘𝅥𝅯𝅘𝅥𝅮𝅘𝅥𝅯 in $\frac{2}{4}$. It may occur at various locations within the measure; musicologists focus on how its position within the bar changed as ragtime evolved over time. In an *untied* syncopation, the pattern starts on either the metrical downbeat or halfway through a measure, therefore occurring entirely in the first or second half of a measure, as in Figure 1(a) and (b). In a *tied* syncopation, the pattern begins either one-quarter or three-quarters of the way through a measure, and therefore either crosses the midpoint of a measure or extends into the following measure, leading to a tie displayed in the notation, as in Figure 1(c) and (d). Music historians and previous studies of ragtime have noted that the untied syncopation was more typical of the early ragtime period of approximately 1890–1901, while the tied syncopation picked up in popularity after the turn of the century [5–7].

3. METHODOLOGY

Our methodology is similar to that used in previous studies, but different enough to warrant some explanation and confirmation that our results align with those of previous work. We begin by preprocessing the RAG-C dataset, with our goal being to identify a set of ragtime pieces sharing a set of basic, consistent properties. We do this by using an established MIR toolkit to identify the time signature and number of parts in each composition. We then use a notation program to automatically quantize the MIDI files and separate the melody from the accompaniment. Because MIDI files do not consistently represent pickup measures (anacruses, or fractional measures at the beginning of a composition), we conduct an experiment to illustrate that they are correctly identified in the dataset. We conclude the preprocessing stage with matching the MIDI files with entries in the RAG-C compendium, which provides useful metadata for each composition, and extracting the rhythms of all the melodic voices.

The RAG-C Dataset. The RAG-C dataset is a corpus of approximately 11,000 ragtime compositions in MIDI format, compiled over time by many enthusiasts of the ragtime genre. In addition to the compositions, the dataset includes a compendium spreadsheet providing metadata for each piece, including title, composer, year (or approximate year) of composition or publication, subgenre within ragtime such as march or two-step, and information about the source of the MIDI file such as the person who played the recording or sequenced the sheet music.

The MIDI file format, having been designed to support communication between electronic music devices, only contains low-level information about the timing of notes in a composition, and therefore MIDI files are more akin to transcriptions of a piece of music rather than a perfect representation of a printed score. A MIDI file is organized into tracks, with each track containing a sequence of *events* specifying when certain notes should be played. Each track also specifies the musical instrument that should play the notes in that track. While it is possible to specify higher-level information such as time or key signatures in a MIDI file, they are not required and are often omitted. Therefore, it is a non-trivial task to extract music-theoretic features from such files [8], especially metrical information. Notes in a MIDI file are only specified as starting and ending at a timestamp given in "ticks" measured from the beginning of the file. It is easiest to infer note durations when a time signature is present in the file and the file is created from a software sequencer, which will ensure that the notes within a MIDI file correspond to a logical metrical grid. When MIDI files are derived from human performances, however, notes which occur simultaneously on the printed page may not match up exactly in terms of ticks due to natural timing variances in performance. *Quantization*, the process of aligning the notes within a MIDI file to a metrical grid, must therefore occur to derive metrical information in such cases.

Melody and Accompaniment Extraction. To address these issues with the RAG-C dataset, we first decided to study only ragtime pieces for solo piano, a decision made in a number of previous studies using the dataset. Within the ragtime piano repertoire, it is common for most of the syncopated action to occur in the right-hand melody part, while the left-hand plays a steady accompaniment. Therefore, isolating the right-hand melody is a clear prerequisite for investigating the syncopated rhythms of ragtime. To accomplish this, we used the `pretty_midi` library [9] to identify MIDI files from the dataset containing exactly two piano tracks in the file. Normally such files contain the melody in one track and the accompaniment in the other; we verified this by calculating the average pitch of the notes in each track and comparing them. The track with the higher average pitch was labeled as the melody, and the other as the accompaniment. To account for files having two piano tracks due to other circumstances (such as two different compositions in one file or a single piano track duplicated twice), we hand-examined all MIDI files where the difference in the average pitches of the two tracks was smaller than one octave to ensure the melody identification algorithm functioned correctly. Furthermore, at this point we ensured that all files remaining had time signatures of either $\frac{2}{4}$, $\frac{4}{4}$, or $\frac{2}{2}$. While ragtime is occasionally found in other time signatures, the vast majority of ragtime music occurs in these meters.

Quantization. Quantizing a MIDI file means aligning the notes in the file to a metrical grid in order to assign natural note durations to each note. To quantize our data, we ran each two-piano-track MIDI file through the MuseScore music notation program [10] and converted each file to its MusicXML representation. MusicXML is a richer format than MIDI that supports more features of common music notation; we use it here specifically so MuseScore can deduce standard note durations and measure boundaries for the MIDI files. We analyzed the resulting MusicXML output files using the `music21` library [11] and discarded any files for which more than 5% of the note onsets in the piece did not align with a 16th note grid. Most ragtime compositions rarely go beyond the 16th note level; we determined that any file with an overabundance of 32nd notes or notes at other onsets in the metrical grid probably was quantized incorrectly.

Title Matching. After all previous steps were completed, we were left with 1991 MIDI files. However, some of these files corresponded to the same ragtime composition, but encoded by different contributors to the RAG-C dataset. To ensure we only had one instance of each composition in our analyses, we used the Levenshtein edit distance to compare each MIDI filename — usually a combination of the composition title and MIDI encoder — against the catalog of composition titles in the RAG-C metadata spreadsheet. Any situation where a filename matched more than title with an edit distance of 5 or less triggered an inspection by hand to assign the proper title. If two or more files matched with a single composition, we kept only the file with the highest quantization percentage; that is, the version with the highest percentage of notes that matched perfectly to the quantization metrical grid. This left us with a final total of 1058 MIDI files in our corpus, each one corresponding to a unique composition.

Accounting for Pickup Measures. Because of the lack of sophisticated metrical information in MIDI files, inconsistencies may arise when processing MIDI files derived from compositions containing a *pickup measure*, that is, an incomplete measure at the beginning of the music. Such music requires special handling as MIDI files do not explicitly store the locations of measure boundaries, and software that assumes such boundaries occur at regular intervals throughout the file will likely incorrectly process a MIDI file containing an incomplete measure. Because we will be investigating syncopation at different points within a measure, it is important that we correctly identify the measure boundaries in such cases.

A common convention is to "pad" a pickup measure with silence at the beginning of a MIDI file, thereby lengthening the first incomplete measure into a complete one. Sometimes this convention is extended to MIDI files that do not have a pickup measure: such files begin with a complete measure of silence. Since any file with silence at the beginning clearly has been padded, we are left with what to do with any unpadded files — we made the decision to treat these files as having full measures throughout, without a pickup measure.

We justify this decision with the following experiment. We identified one particular contributor to the RAG-C dataset who chose to always encode MIDI files with silence at the beginning: a padded partial measure of silence for compositions beginning with a pickup, or a complete measure of silence for compositions without a pickup. This individual encoded 104 compositions in the dataset, and 26 of them began with a partial measure of silence, versus 78 with a complete measure of silence. Because 25% of this particular contributor's files begin with a pickup measure, we would expect that this proportion would hold in the remainder of the corpus as well. Of the 1,991 MIDI files remaining after the quantization step, 1887 of them do not come from the contributor in question. Of the 1887, 539 begin with a fractional measure of silence and 1348 begin with a complete measure of silence or no silence. Because 539 out of 1887 is approximately 28.6%, it is reasonable to assume that the files with no silence at the beginning correspond to compositions with no pickup measure.

Binary Onset Patterns. The last remaining step in preprocessing the RAG-C dataset is to identify the rhythms of the melody part. Previous studies of ragtime syncopation used the convention of *binary onset patterns* to represent rhythms. These patterns are sequences of ones and zeros where a one represents the onset of a note and a zero represents a continuation of a note or a rest. These patterns can be computed at different levels of metrical granularity from a score. For instance, the binary onset pattern for a $\frac{2}{4}$ measure of four eighth notes would be "`10101010`" computed at the 16th-note level, but "`1111`" computed at the eighth note level.

Our dataset contains ragtime compositions in three different time signatures, namely $\frac{2}{4}$ (810 pieces), $\frac{4}{4}$ (214 pieces), and $\frac{2}{2}$ (34 pieces). We chose to compute binary onset patterns at the sixteenth note granularity for $\frac{2}{4}$ compositions, and at the eighth note granularity for $\frac{4}{4}$ and $\frac{2}{2}$ pieces. The basis for this decision was the observation that the use of sixteenth notes differed between pieces notated in the three time signatures, verified by the following experiment. We define the *sixteenth note density* for a measure of music as the proportion of the *weak* sixteenth note beats in a measure (that is, onsets 2, 4, 6, and 8 in $\frac{2}{4}$) that contain at least one note onset. We define the *eighth note density* for a measure similarly. We then computed the distribution of densities of sixteenth notes in all three time signatures, and eighth note densities in $\frac{4}{4}$ and $\frac{2}{2}$. We observed that our initial choices of appropriate granularities for binary onset patterns — corresponding to the first three plots in Figure 2 — fall in similar ranges, while the last two plots do not, justifying our mixed use of sixteenth and eighth note binary onset patterns. However, Figure 2 does show a number of outliers in the $\frac{4}{4}$ and $\frac{2}{2}$ sixteenth note density plots, falling more in the range of the first three plots, indicating

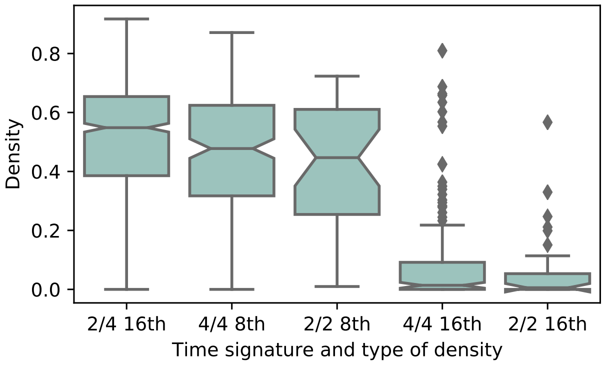


Figure 2. Box plots illustrating the distribution of 8th and 16th note densities in the various time signatures.

that in future experiments, we should consider analyzing those particular pieces at the sixteenth note level.

4. EXPERIMENTS

In this section we describe a number of analytical experiments we conducted to extract information from our corpus about the way syncopation is used in ragtime music.

4.1 Exploring the 121 Pattern

Ragtime scholars hypothesized *untied* syncopations were the predominant form of syncopated pattern found in early ragtime compositions from approximately 1890 to the turn of the century, while *tied* syncopations did not become common until around 1902 in the late ragtime period [6,7]. This hypothesis was confirmed by Volk and De Haas [1]; we replicate their experiment due to differing methodologies for processing the RAG-C dataset. These differences in selecting and quantizing MIDI files naturally produce a slightly different corpus with which we are working, and therefore replicating earlier work shows that these results are invariant with a well-rounded corpus and lends credence to our extended results that build on earlier studies.

In this experiment, we compared the frequency of use of the 121 tied and untied patterns in ragtime compositions from three eras: first, we compared the early ragtime period of 1890–1901 (110 pieces) with the late ragtime period of 1902–1919 (582 pieces), and then compared the entire ragtime period of 1890–1919 (692 pieces) with the modern period of 1920 to the present (362 pieces). The 121 patterns were found by looking at the binary onset patterns that were collected earlier and counting the number of times each variety of syncopation appeared in a composition. It is possible for multiple 121 syncopations to appear in a single measure, as each syncopation covers only part of a measure. To account for differing lengths of pieces, we divided the total tallies by the number of measures in each composition, resulting in *frequency per measure*.

Overall, these results are similar to those of Volk and De Haas, taking into account some variation for differing pieces selected from the RAG-C dataset. In particular, we confirm that the number of tied patterns doubled between the early and late ragtime eras, and the number of untied patterns decreased between the eras as well. The left two

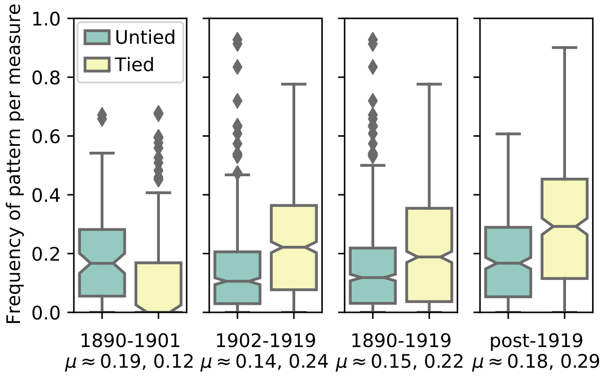


Figure 3. Box plots illustrating the distribution of frequencies of 121 patterns per measure, comparing different eras of ragtime. Means (μ) are shown below each plot.

plots in Figure 3 illustrate this. Wilcoxon rank-sum tests confirm (untied: $p < 0.001$, tied: $p \ll 0.001$) that the differences between the eras are statistically significant given the null hypothesis that the distributions are identical.

The right two plots in Figure 3 can be analyzed in a similar fashion, and illustrate that the use of both types of syncopation climbed after the end of the ragtime era. Wilcoxon rank-sum tests again confirm (untied: $p \ll 0.001$, tied: $p \ll 0.001$) that these increases are statistically significant.

The Big Three. Ragtime scholars agree that three ragtime composers stand out from the rest in terms of best exemplifying the ragtime genre: Scott Joplin (1867 or 1868–1917), James Scott (1885–1938), and Joseph Lamb (1887–1960) [12–14]. These composers are well-represented in the RAG-C dataset, and it is instructive to examine if they used syncopation patterns differently than each other or compared to other ragtime composers.

When comparing the big three among themselves, small differences in usage of the 121 pattern emerge but nothing that cannot be attributed to chance. However, statistically significant differences are evident when comparing the output of the big three against other ragtime composers. In particular, even when controlling for era, the big three composers used more 121 patterns than other composers. In this experiment, we isolated the output of the big three composers during the late ragtime era of 1902–1919 (70 pieces), and compared their compositions against those of the remaining composers from the same era (512 pieces). Wilcoxon rank-sum tests confirm that the differences in the frequency of use of both untied ($p < 0.0001$) and tied ($p < 0.0001$) syncopations are different between the big three and the remaining composers. Figure 4 illustrates how the big three used, on average, almost 70% more untied 121 syncopations and almost 50% more tied 121 syncopations.

4.2 Analyzing other syncopated patterns

Analyzing frequencies of the 121 pattern is instructive to verify certain hypotheses put forth by musicologists. However, it is not a given that this pattern or its variants are necessarily the most prevalent patterns found in ragtime. In

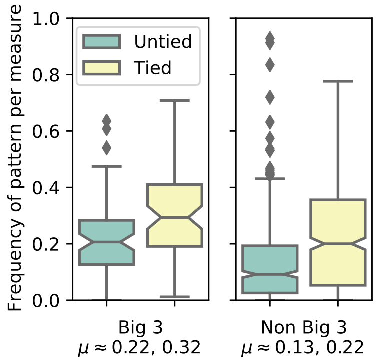


Figure 4. Box plots illustrating the distribution of frequencies of 121 patterns per measure, comparing compositions by the "big three" ragtime composers in the late ragtime era versus all other compositions from that same era.

this section, we expand our analysis to explore all possible binary onset patterns and measure the amount of syncopation present in each pattern.

We follow the model of Koops, et. al. [2] and use the Longuet-Higgins and Lee [15] metric, grounded in aural perception of rhythm [16], to quantify the amount of syncopation present in a measure based on its binary onset pattern. The LHL metric is zero for a measure with no syncopation and increases with each instance of a note onset (`1`) occurring on a weak beat followed by no onset (`0`) on the immediately-following (strong) beat. The increases are larger for syncopations crossing more significant divisions of the measure. For example, the binary onset pattern `01010101` contains three syncopations (instances of `10`). The syncopation in the middle that crosses the midpoint of the measure has an LHL value of 2, and the two syncopations on either side have values of 1 since they cross weaker divisions of the measure. Therefore, this measure by itself has an LHL score of 4. If this measure were followed by no onsets on the downbeat of the next measure, the LHL value would increase by 3 for the additional syncopation crossing the barline, for a total LHL value of 7.

In performing the following experiments, we calculated the LHL values for each measure of the melody in the corpus using the binary onset patterns computed earlier. Our corpus of 1,058 pieces contained 140,856 measures of music, with the average LHL value for a measure being 1.17, with a standard deviation of 1.35. However, only 77,012 of the measures ($\approx 55\%$) contained any syncopation at all (LHL > 0). If we only consider measures with LHL > 0, the average LHL value becomes 2.14, with a standard deviation of 1.12.

We note that our methodology for computing and interpreting binary onset patterns differs enough here from Koops, et. al. [2] to warrant explanation. In their work, binary onset patterns may be up to 16 bits in length, corresponding to measures in $\frac{4}{4}$ or $\frac{2}{2}$ analyzed at the 16th note level, allowing for the possibility of having LHL values for a single measure of music as high as 15. Here, all our binary onset patterns are of length 8, due to always analyzing measures in $\frac{4}{4}$ or $\frac{2}{2}$ at the eighth note level and measures in $\frac{2}{4}$ at the 16th note level. Therefore, it is difficult to draw

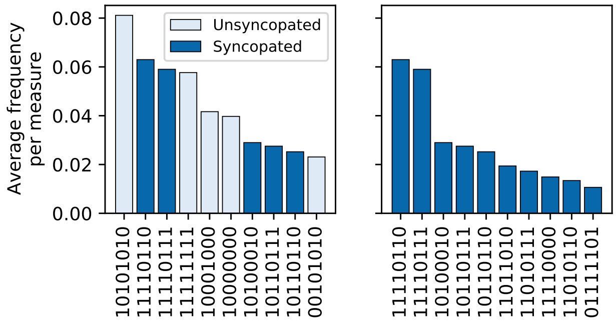


Figure 5. Left: The ten most frequent binary onset patterns overall, differentiating between unsyncopated patterns (LHL > 0) and syncopated patterns (LHL > 0). Right: Then ten most frequent syncopated patterns.

direct numerical equivalences between our studies.

Overall patterns. We first examine the frequencies of all possible binary onset patterns in the corpus. To ensure that the varying lengths of the compositions in the corpus did not affect our results, we divided the number of times each binary onset pattern appeared in a composition by the number of measures in that composition, thereby obtaining the *frequency per measure* for each pattern. We then averaged the frequencies across all pieces in the corpus. The top ten patterns overall, and the top ten with an LHL score greater than 0 are shown in Figure 5. It is noteworthy that for a genre identified with such high levels of syncopation, there are many common non-syncopated patterns. In particular, patterns 1, 4, 5, 6, and 10 do not contain any syncopation.

In analyzing the right side of Figure 5, we note the presence of a tied 121 pattern (`1101`) in the middle of patterns 1, 2, 4, and 5, along with untied 121 patterns at the beginning of patterns 6, 7, and 9, and at the end of pattern 10. Additionally, pattern 3 is the 121 pattern in augmented form. Only pattern 8 is not connected with the 121 figure.

Patterns by era. We conducted an experiment to determine whether composers of different eras used certain types of binary onset patterns differently. Using our earlier grouping of compositions into the early ragtime era, the late ragtime era, and the modern era, we computed the most popular rhythms in each group, which can be seen in Figure 6; we note that popular patterns in some eras become unpopular in others. We used three Wilcoxon signed-rank tests to compare pairs of eras, using all 256 possible binary onset patterns in each test. The results give us weak statistical significance at the $\alpha = 0.05$ level using the Šidák correction to account for the multiple comparisons, suggesting that composers chose rhythmic patterns differently in the three eras ($p \approx 0.01$ for the early versus late eras, $p < 0.001$ for the early versus modern eras, and $p \approx 0.01$ for the late versus modern eras).

4.3 Transitions Between Patterns

The overall sound of a piece of music depends not only on the contents of the individual measures of music but also on the choice of which measures follow other measures. Ragtime is no exception, and we hypothesize that there are

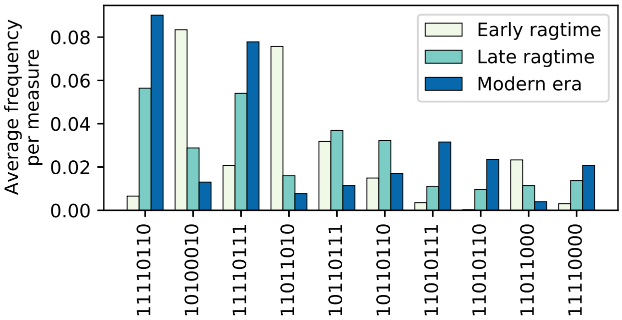
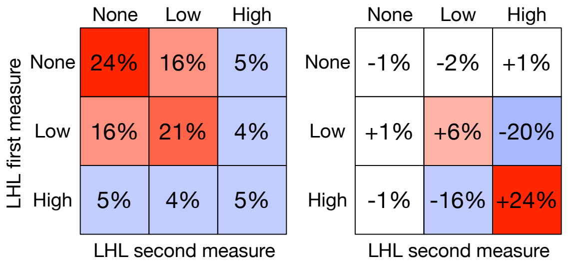


Figure 6. The most frequent binary onset patterns, segmented by era.



(a) Raw transition probabilities (b) Deviation from expected

Figure 7. (a) The joint transition probabilities for observing of the nine possible LHL transition pairs. (b) The deviation from the expected probability if there were no correlation between the LHL value in one measure and the next. The four highlighted deviations in (b) are statistically significant.

relationships between the syncopated patterns in consecutive measures. Specifically, we propose the question of whether the degree of syncopation in a measure of music is related to the degree of syncopation in the surrounding measures.

We chose to answer this question by examining all consecutive pairs of measures of music in the corpus and computing their LHL values separately for the first measure and the second measure of the pair. Recalling that the average LHL value for a syncopated measure of music was approximately 2.14, we binned the LHL values according to having a *high* amount of syncopation (LHL $\geq$ 3), a *low* amount of syncopation (LHL = 1 or 2), or *no* syncopation (LHL = 0). For each piece of music, we computed the frequencies of each of the nine possible LHL transition pairs, normalizing for the number of measures in each composition, and averaged the frequencies across all compositions. The results are displayed as joint probabilities in a heatmap in Figure 7(a). We observe that transitions between consecutive measures with no syncopation are extremely common, while any transition involving a high amount of syncopation is uncommon.

These joint probabilities, however, do not tell the whole story. Because the distribution of LHL values is highly skewed towards the smaller values, it is worthwhile to test if any of these LHL transition pairs occur with certain tendencies due to chance, or due to deliberate choices on the composer's part. For example, given that almost half of the measures of music in the corpus do not contain any syncopation, should we be surprised that 24% of the total transitions are between two measures without syncopation? We can answer this question with a final experiment. We compared the LHL transition frequencies obtained from the corpus against corresponding frequencies that would be obtained if one were to randomly reorder the measures in each piece of music. Specifically, for each composition, we generated 1000 random reorderings of the measures, computed the LHL transitions for every pair of consecutive measures, averaged them across the 1000 reorderings, and then proceeded as we did earlier with normalizing and averaging across all compositions. These results, illustrated in Figure 7(b), confirm that a number of the LHL transitions occur significantly more or less frequently than would be expected under a reordering of measures of the compositions.

We used nine individual binomial tests to compare the the true LHL transition frequencies to the expected frequencies under the null hypothesis that measure transitions resemble those done randomly. At a significance level of $\alpha = 0.05$, taking into account the Šidák correction for multiple comparisons, the four highlighted transitions in Figure 7(b) are statistically significant (all with $p \ll 0.0001$). This tells us, for instance, that even though it is overall rare to find consecutive measures with high amounts of syncopation in a composition, this phenomenon still occurs more often than would be expected if a composer were introducing syncopation at random.

5. CONCLUSION AND FUTURE WORK

In this study, we used new methods to extend a number of previous analyses of ragtime syncopation to confirm existing musicological hypotheses and present new ones. Specifically, we confirmed the earlier finding that the 121 syncopation idiom varies in use between ragtime eras, even when using a different preprocessing scheme for the RAG-C dataset. We demonstrated a new finding that the "big three" ragtime composers also employed this syncopation pattern more often than their contemporaries did. We illustrated how other rhythmic patterns evolved over time and revealed different frequency distributions. Lastly, we displayed novel results showing statistically significant differences in the way the amount of syncopation changes between consecutive measures in a ragtime composition.

In future work, we plan to continue to study the use of syncopation in ragtime, specifically towards uncovering more information about varying musical parameters (rhythmic, melodic, or harmonic) between measures. We believe it will be useful to expand the ideas in this paper to other musical genres as well. Previous research has successfully used information about rhythmic patterns to assist in genre classification [17–19], and we imagine the data presented here could be useful in such circumstances. We also hypothesize that algorithmic composition techniques that rely on probabilistic techniques for rhythm generation [4, 20] might be improved by considering how musical parameters like rhythm change from measure to measure as well as within a single measure.

6. REFERENCES

[1] A. Volk and W. B. de Haas, "A corpus-based study on ragtime syncopation," in *Proc. of the 14th International Society for Music Information Retrieval Conference*, 2013, pp. 163–168.

[2] H. V. Koops, A. Volk, and W. B. de Haas, "Corpus-based rhythmic pattern analysis of ragtime syncopation," in *Proc. of the 16th International Society for Music Information Retrieval Conference*, 2015, pp. 483–489.

[3] D. Odekerken, A. Volk, and H. V. Koops, "Rhythmic patterns in ragtime and jazz," in *Proc. of the 7th International Workshop on Folk Music Analysis*, 2017, pp. 44–49.

[4] J. Michelson, H. Xu, and P. B. Kirlin, "Probabilistic generation of ragtime music from classical melodies," in *Proc. of the 6th International Conference on Mathematics and Computation in Music*. Springer International Publishing, 2017, pp. 350–360.

[5] I. Harer, "Defining ragtime music: Historical and typological research," *Studia Musicologica Academiae Scientiarum Hungaricae*, vol. 38, no. 3–4, pp. 409–415, 1997.

[6] E. A. Berlin, "Ragtime," in *Grove Music Online*. Oxford University Press, 2013. [Online]. Available: https://doi.org/10.1093/gmo/9781561592630.article.A2252241

[7] J. E. Hasse, "Ragtime: From the top," in *Ragtime: Its History, Composers, and Music*, J. E. Hasse, Ed. New York: Shirmer Books, 1985, pp. 1–39.

[8] C. Raffel and D. P. W. Ellis, "Extracting ground-truth information from MIDI files: A MIDIfesto," in *Proc. of the 17th International Society for Music Information Retrieval Conference*, 2016, pp. 796–802.

[9] ——, "Intuitive analysis, creation and manipulation of MIDI data with `pretty_midi`," in *Proc. of the 15th International Conference on Music Information Retrieval, Late Breaking and Demo Papers*, 2014.

[10] MuseScore BVBA, "MuseScore," version 3.4.2. Software. http://www.musescore.org.

[11] M. S. Cuthbert and C. Ariza, "`music21`: A toolkit for computer-aided musicology and symbolic music data," in *Proc. of the 11th International Society for Music Information Retrieval Conference*, 2010, pp. 637–642.

[12] R. Blesh and H. Janis, *They All Played Ragtime*, 4th ed. New York: Oak Publications, 1971.

[13] W. M. McNally, "Ragtime then and now: Composers and audiences from the ragtime era to the ragtime revival," Ph.D. dissertation, The City University of New York, 2015.

[14] J. Magee, "Ragtime and early jazz," in *The Cambridge History of American Music*, D. Nicholls, Ed. Cambridge University Press, 1998, pp. 388–417.

[15] H. C. Longuet-Higgins and C. S. Lee, "The rhythmic interpretation of monophonic music," *Music Perception*, vol. 1, no. 4, pp. 424–441, 1984.

[16] ——, "The perception of musical rhythms," *Perception*, vol. 11, no. 2, pp. 115–128, 1982.

[17] S. Dixon, F. Gouyon, and G. Widmer, "Towards characterisation of music via rhythmic patterns," in *Proc. of the 5th International Conference on Music Information Retrieval*, 2004, pp. 509–516.

[18] D. C. Correa, L. d. F. Costa, and J. H. Saito, "Tracking the beat: Classification of music genres and synthesis of rhythms," in *Proc. of the 17th International Conference on Systems, Signals and Image Processing*, 2010, pp. 280–283.

[19] T. M. Esparza, J. P. Bello, and E. J. Humphrey, "From genre classification to rhythm similarity: Computational and musicological insights," *Journal of New Music Research*, vol. 44, no. 1, pp. 39–57, 2015.

[20] J. D. Fernández and F. Vico, "AI methods in algorithmic composition: A comprehensive survey," *Journal of Artificial Intelligence Research*, vol. 48, pp. 513–582, 2013.