

# A CORPUS-BASED STUDY ON RAGTIME SYNCOPATION

Anja Volk  
Utrecht University  
A.Volk@uu.nl

W. Bas de Haas  
Utrecht University  
W.B.deHaas@uu.nl

## ABSTRACT

This paper presents a corpus-based study on syncopation patterns in ragtime. We discuss open questions on the ragtime genre and the potential of computational tools in addressing these questions, contributing to the fields of Musicology and Music Information Retrieval (MIR), and giving back to the ragtime enthusiasts community. We introduce the RAG-collection of around 11000 ragtime MIDI files collected, organised, and distributed by many ragtime lovers around the world. The collection is accompanied by a compendium, providing useful metadata on ragtime compositions. Using this collection and the compendium, we investigate syncopation patterns in ragtime melodies, for which we tailored a melody extraction algorithm. We test and confirm musicological hypotheses about the occurrence of syncopation patterns that are considered typical for ragtime on the extracted melodies. Thus, the paper presents a first step towards modelling typical characteristics of the ragtime genre, which is an important means for enabling automatic genre classification.

## 1. INTRODUCTION

This paper presents a corpus-based study on ragtime syncopation using a collection of around 11000 ragtime MIDI files. Corpus-based studies have the potential to quantitatively answer research questions in musicology using a data-rich approach, thus providing a complementary strategy to methods in Musicology. Moreover, they provide examples of how Music Information Retrieval (MIR) methods can successfully contribute to other research areas, which has recently been discussed as an important issue of interdisciplinarity in MIR [1, 15]. In this paper we argue that corpus-based studies also contribute to important research questions within MIR, such as similarity estimation and genre classification.

Mauch and Dixon have carried out a corpus-based study of bar-length drum patterns in a collection of 48000 MIDI files gathered from the Internet, motivated by the observation that “comparatively little work in MIR has quantitatively examined rhythms in symbolic data” [11, p. 164]. The paper argues that the outcome of such a study might be of direct interest for musicians and musicologists. Con-

clusions from the detected repeated drum patterns in this collection are drawn on a general level, such as about the difference between language and music with regard to repetition. In this paper we argue that the outcome of a corpus-based study is even more meaningful to musicians and musicologists, if the questions addressed are meaningful for the specific corpus investigated. Hence, we demonstrate what questions on ragtimes exist among musicologists and music lovers, and why a corpus-based study on ragtimes is therefore of vital interest to musicologists, music lovers and MIR researchers.

In this paper we introduce a corpus of around 11000 ragtime MIDI files as the dataset for our study on syncopation patterns, the RAG<sup>1</sup>-collection. The MIDI files of this collection have been produced, organised and collected by many ragtime musicians and lovers all around the world and shared via the Internet. The collection has grown over many years, and an accompanying Ragtime Compendium has been created by Michael Mathew and colleagues with metadata about the corresponding ragtime compositions (such as title, year, composer, original publisher, website of the MIDI file) [10]. The RAG-collection is the product of a shared effort of ragtime enthusiasts and hence presents a heterogeneous collection of MIDI files of very different quality due to their different origins (such as original piano rolls that have later been translated into MIDI format, actual performances of ragtimes or encodings of the notated score). In this paper we undertake first steps to use the RAG-collection as a basis for investigating characteristics of the ragtime genre employing MIR methods.

The contribution of the paper is threefold. First, we give an overview over open questions on ragtime as a genre relevant to musicologists, MIR researchers, and ragtime lovers; we argue how MIR can productively contribute to answering these questions with a corpus-based study. Second, we investigate syncopation patterns on the RAG-collection as a means of understanding rhythmic characteristics of this genre. Specifically, we test and confirm musicological hypotheses about the occurrence of so-called *tied* and *untied* syncopation patterns in ragtimes during different time periods of *the ragtime era* (1890-1919). Furthermore we show that these patterns remained important for ragtimes in the period 1920-2012 and have been used even more extensively. Third, we tailor MIR tools for melody finding to the ragtime genre. By investigating syncopation patterns in ragtimes using a data-rich approach, we undertake a first step towards modelling typical characteristics of

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page.

© 2013 International Society for Music Information Retrieval.

<sup>1</sup> Ragtime Admirers Group

the ragtime genre, which is an important means for automatic genre classification based on similarity, contributing to a core topic in MIR.

## 2. SIGNIFICANCE OF A CORPUS-BASED STUDY ON RAGTIMES FOR MUSICOLOGY AND MIR

In this section we argue that a corpus-based study on ragtimes contributes to important issues in both MIR and Musicology regarding the question on what constitutes the genre of ragtime.

### 2.1 Open questions on ragtime as a genre

While most people associate *ragtime* with the piano music by Scott Joplin, James Scott and Joseph Lamb, ragtime music in fact comprises a wide range of vocal, instrumental, improvised and composed music with lots of different musical influences (such as difference dance music styles like march, two-step, habaneras, or tangos). Ragtime is considered a “hybrid from folk and written cultures” [5, p. 65]; as a consequence, ragtime is both investigated in the context of “folk”, “popular” and “art” music [4, p. 142]. As of today, ragtimes present “thorny problems” to musical scholars [4], one of those crucial problems concerns the very identity of what is ragtime: “A major question that emerged in the course of this musical survey was how to determine what music could reasonably be considered ragtime” [2, p. xviii]. As we will argue later, this question is not only of scholarly interest, but also concerns the vital community of ragtime lovers all over the world.

One of the main characterizations of ragtimes is related to syncopation, as Edward Berlin points out: “At the core of the contemporary understanding of ragtime ... was syncopation. The question ‘What is ragtime?’ was asked throughout the period, and almost invariably explanations included a statement about syncopation” [2, p. 11]. However, there are ragtimes that do not contain syncopation, and there exist lots of syncopated music that is not considered ragtime. Hence, the role of syncopation for the ragtime genre needs still to be more specified, and additionally other features need to be described that contribute to ragtime as a genre. One example of investigating the specific role of syncopation in ragtime is that musical scholars have determined typical syncopation patterns in ragtime alongside with hypotheses on what patterns are most typical for a certain period of ragtime [2, 7], which we will address in this paper using computational approaches.

Closely related to the question of what constitutes a ragtime is the question about the different origins or musical sources that contributed or are still contributing to ragtime, and what subgroups within this broad genre can be identified. There exist different theories as to what has contributed to shaping the genre of ragtime, such as coon-songs, cakewalk, two-steps, or dance music of the Caribbean or South America (such as danzas, habaneras, and tangos) and music characteristics of African American origins. With regard to piano ragtimes, Berlin distinguishes at least three different subgroups: “piano renditions of ragtime songs; ‘ragged’ versions of pre-existing

unsyncopated music; and original, dance-oriented ragtime compositions” [2, p. 63]. Furthermore he lists patriotic and folk tunes as well as classical music pieces as common musical sources that by *being ragged* have become members of the ragtime repertoire. In the U.S. many local Ragtime societies and festivals exist, and attempts have been made to distinguish stylistically different types of ragtimes dependent on their local origin – however, this has not yet been achieved according to Berlin [3].

Moreover, Berlin discusses that what has been considered ragtime by listeners in the past and present, has also been shaped by socio-cultural aspects: “My original search for a working definition of ragtime was almost dwarfed as I became immersed in the conflicts of its day, concerning ragtime’s origins, racial content, relevance to American music, innovative features, potential for ‘artistic’ development, effects on cultural, moral, and physical well-being and the like” [3, p. xviii].

### 2.2 The significance of a corpus-based investigation of ragtimes for musicology

An investigation of ragtimes using a large digital corpus comprising ragtime pieces from the 1890s to present is necessary in order to overcome current biases about this genre due to a rather restricted perspective that comes from the musical pieces that have been mainly investigated so far. The years of roughly 1890s–1919 are considered as *the Ragtime era* when the genre was established, and for the general public a resurrection of the genre took place following the film “The Sting” in the 1970s. However, in fact ragtime music has been a lively music tradition also in the years in between. Edward Berlin argued that published anthologies of ragtimes over-represent “the best-known composers” (Joplin, Scott, Lamb) from the Ragtime era, and that therefore in ragtime research the stylistic traits of their music has overweighted stylistic traits of other ragtime compositions [2, p. 72]. Therefore, in order to get a more balanced view of the ragtime genre as a whole, he has carried out by hand a study of a random selection of ragtimes at the Lincoln Center. Studying a large digital ragtime corpus with computational tools is the logical next step of broadening the perspective on ragtime. Moreover, attempts to define possible subgroups in the broad range of existing ragtimes, provide a natural setting for research on genre using a computational data-rich approach.

Investigating rhythmic patterns in ragtime also contributes to research on variation in music. Variation is a fundamental trait in music and is linked to the experience of similarity in music, a central concept in MIR [14]. Most often, variation has been studied in the context of Variation sets in Western classical music, where one musical idea is widely exploited by the composer. However, as a means of establishing similarity in music, it seems more appropriate to study the relation between variation patterns and style, as suggested by Meyer [12, p. 3]: “Style is a replication of patterning”. For ragtime there exist musicological hypotheses on what patterns are characteristic for the genre; the role of these patterns can be tested with computational

methods. In this paper we test hypotheses about the occurrence of a typical syncopation pattern in ragtime, which has been dubbed as *Leitmotiv*<sup>2</sup> of ragtime by musicologists [6]. We examine how often and in which variations with respect to the location within the bar, this pattern occurs within the RAG-collection.

### 2.3 The significance of a corpus-based investigation of ragtimes for MIR and ragtime lovers

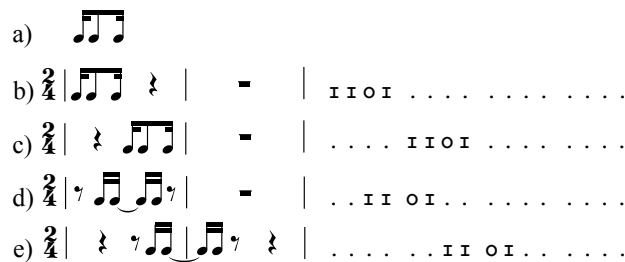
The question on how to determine with computational methods the genre of a given digitized musical piece lies at the heart of MIR research. At the same time, it is all too often very unclear what a reasonable argument would be to classify a certain musical piece into a given genre category. The genre of ragtime provides an interesting case study of linking rhythmic and harmonic patterns that are discussed as being constituents of the genre to the occurrence of these patterns within a large digitized corpus of ragtime pieces.

Moreover, the question of what constitutes a ragtime is not only interesting from a scholarly perspective, but is of interest to a rather large international ragtime fan community (with ragtime festivals and societies<sup>3</sup> existing) that shares lots of digital collections on the Internet: “the ragtimers . . . have clubs, magazines, and an international following. The subject owes an enormous amount to these individual collectors and dedicated players who have helped to preserve the repertoire and keep it alive. New rags are being composed and played all the time” [5, p. 65].

Piano rolls belong to the first sources that preserved ragtimes before they appeared as printed scores in sheet music. Many of these piano rolls have been later preserved as MIDI files, and along with many other compositions for which MIDI files have been created and shared via the Internet. We have now a collection of around 11000 MIDI files that have been recognized as ragtime by ragtime lovers. It is of vital interest, also to this community of ragtime fans, whether a given digital file might be considered as ragtime or into what subcategory of ragtime it might belong: “A discussion on what constitutes ragtime could fill several web pages” [10]. While socio-cultural aspects certainly contribute to constituting the ragtime genre, content-based MIR-methods can assist in investigating in how far musical characteristics contribute to the the identity of ragtime. Moreover, MIR-methods can be used to identify variants of the same ragtime composition in different MIDI file renditions existing on the Internet. Content-based MIR-methods hence would be of great use for this community, since metadata on ragtimes is rather problematic and not very reliable (i.e. ragtime composers have used different pseudonyms for different genres [10]). Hence, ragtimes provide an interesting application area for content-based methods on similarity in MIR.

<sup>2</sup> by metaphorically extending the original concept of *Leitmotiv* from recurring motifs that represent a certain person, idea etc. to recurring motifs that represent a genre

<sup>3</sup> e.g. <http://www.westcoastragtime.com>; <http://indianarag.org>; <http://www.bohemragtime.com/en/act.html>



**Figure 1.** (a) shows the syncopation pattern dubbed *I2I* in [7] and its different shifts within the bar, leading to *untied* syncopations in (b) and (c), and *tied* syncopations in (d) and (e). Behind the pattern scores we display the patterns used for matching: I denotes an onset, O denotes no onset, and . denotes a wild-card that matches everything. Every position represents sixteenth note position within a bar.

### 3. THE RAG-COLLECTION

The RAG-collection currently contains 11591 MIDI files of ragtimes gathered from the Internet. The Ragtime Compendium [10] is an accompanying database on the identification of individual ragtime compositions (year, title, composer, publisher) and whether or not a MIDI file exists. The Compendium lists around 15000 identified compositions (each entry corresponds to a unique ragtime composition), out of these 5600 pieces have MIDI files. Hence, the 11591 MIDI files contain lots of different MIDI renditions for the same composition. The compendium “has evolved over several years, starting from very humble beginnings and now, thanks to the help of many ragtime enthusiasts, it is probably the most complete listing that exists” [10].

### 4. A CORPUS-BASED STUDY ON SYNCOPATION

For the characterization of musical pieces as ragtimes, syncopation plays a major role, as described in section 2. To be able to distinguish between pieces that contain syncopations and that belong to the ragtime genre or not, more detailed investigations of syncopation patterns are necessary. Are ragtimes characterized by a higher amount of syncopation patterns than other genres, and/or are there specific syncopation patterns prevalent in ragtimes more than in other music styles? Musicologists [2, 7] discuss a number of syncopation patterns typically occurring in ragtimes of *the ragtime era* (1890s–1919), along with hypotheses on what patterns occur most often in early (until 1901) or late ragtimes (1902–1919) based on an extensive study of ragtime pieces from this era by hand.

In this paper we are testing these hypotheses about typical syncopation patterns with a computational approach, as a first step of exploring the role of syncopation in defining ragtimes. Specifically, we test Berlin’s hypothesis that so-called *untied* syncopations are a typical characteristic trait of early ragtimes, while *tied* syncopations have been introduced in the late ragtimes starting roughly from 1902. Hence, many early ragtimes use *untied* syncopations exclusively, while tied syncopations became the predominant type of syncopation later [2, p. 128]. The difference between tied and untied syncopation is discussed with the

syncopation pattern shown in Figure 1(a), which has been called the *Leitmotiv* of ragtime music [6] and called pattern 121 in [7]. Shifting this pattern along the bar leads to either untied syncopation (Figures 1(b) and (c)) or tied syncopation (Figures 1(d) and (e)).

We were unable to find musicological hypotheses about the role of these syncopation patterns in ragtimes that have been composed after 1919 (following the ragtime era). Since the amount of ragtime pieces composed dropped significantly after 1919, one might hypothesize that by then the consolidation of the ragtime style has taken place. Ragtimes following this consolidation might therefore exhibit similar amounts of untied and tied syncopation as in the second half of the ragtime era (1902–1919). However, since the extensive use of syncopation patterns was new to the general musical audience at the beginning of the 20th century, one might also hypothesize that the amount of syncopation patterns increases over time due to a common habituation to syncopations. For instance, an increased use of syncopation in American popular music in general has been concluded from an empirical study in [9], using 437 short random samples from pieces dating back from 1890 to 1939. Yet a third hypothesis arises from Berlin’s remarks on the erosion of ragtime’s “most distinctive qualities” [2, p. 147] that started in the 1910s. The increased use of dotted rhythms observed in late ragtimes “reduced the need for syncopation in ragtime” [2, p. 148]. From this perspective, a decreased use of syncopation might be hypothesized for the following years. We take advantage of the RAG-collection containing ragtime pieces until 2012 in order to test the role of these syncopation patterns in ragtimes composed after the ragtime era.

## 4.1 Data preparation

For testing hypotheses on syncopation patterns, we first need to pre-process the MIDI files of the RAG-collection.

### 4.1.1 Melody finding

Within the ragtime genre most of the interesting rhythmical material is in the melody, hence in the case of a piano piece, this mainly concerns the right hand part of the piece. The accompaniment in the left hand is characterised by stable rhythmical patterns that follow the beat (also dubbed *oompah* bass line). Since the melody is not annotated explicitly in the MIDI files of the RAG-collection, we need to employ automatic melody extraction in order to split the melody from the accompaniment. We use a version of the *skyline* algorithm for melody finding [16]; this algorithm takes the highest sounding note when multiple notes sound simultaneously. To overcome that highest notes from the left hand are classified wrongly as part of the melody at sections where there is no melody, we set a lower limit: all notes below the middle C are classified as accompaniment.

We evaluated our skyline implementation on a small selection of 435 songs of the RAG-collection. These songs were selected automatically based on whether they were quantised, and had exactly two tracks. We ensured that there were no duplicate files or tracks. In most cases the

	P	R	F
Melody extraction	.930	.989	.958
Melody extraction w. dip-detection	.972	.984	.978

**Table 1.** The Precision, Recall, and F-measure of the melody extraction with and without *dip-detection*, evaluated on a 435-piece subset of the RAG-collection.

first track was the melody track, but in a few cases the melody and accompaniment track were reversed. Hence, we checked whether both the lowest and the highest notes in the melody were higher than the lowest and highest notes in the accompaniment. If this was not the case, we examined the file by hand. When the two tracks did not clearly reflect a melody and accompaniment, we removed it from the selection. Next, we performed a melody extraction experiment by creating one track containing all notes, performing the melody finding, and comparing the result to the original melody track (our ground-truth melody). In some cases the ground-truth melody track contained chords, since we are mainly interested in rhythm, and did not tailor our skyline implementation for retrieving chords in the melody track, we removed all but the highest chord notes.

We compared the extracted melody to our ground-truth melody by calculating the precision, recall and F-measure for every piece, and averaged these numbers over the total selection of 435 songs. The results are displayed on the first row of Table 1, and can be considered good. The high recall indicates that almost all notes of the original melody are retrieved. The high precision shows in turn that we do not include a lot of non-melody notes into our automatically extracted melody.

Although the melody extraction could be considered good, manual inspection of some extracted melodies revealed that accompaniment notes were classified as melody notes typically at occasions where syncopation in the right hand occurs. In absolute numbers there were not many of these spurious melody notes. However, since they broke some of the syncopation patterns that we are interested in (see Figure 1), we extended our melody finding algorithm with a *dip-detection*: most of the misclassifications can be recognised by a large interval down followed by a large interval up. Therefore, after performing the skyline algorithm, we removed notes that were characterised by an interval down greater than 9 semitones followed by an interval up greater than 9 semitones.

We evaluate the melody extraction with dip-detection on the same subset of the RAG-collection, the results are displayed in the second row of Table 1. They show that the precision improves, while the recall hardly changes. Overall the skyline algorithm with dip-detection yields a near-perfect result on this subset of the RAG-collection.

### 4.1.2 Quantisation

To be able to analyse the rhythmical patterns in the ragtime melodies, the MIDI data needs to be quantised appropriately. Although most ragtime performers claim that ragtime should be played with *straight* eighth notes, we found

a large number of MIDI files that contained ragtimes performed with *swing*. Swing refers to a characteristic long-short subdivision of the beat that is generally considered a crucial aspect contributing to the quality of a jazz or pop performance (see [8] for further information). To be able to capture and represent all of these different subdivisions of the beat properly, we shift all onsets onto a grid of 12 equally spaced divisions per quarter note. To estimate how well a file can be aligned to the metrical grid, we store the quantisation deviation for every note (in MIDI ticks), and calculate the average. Because MIDI files can have different time resolutions, we divide the average deviation by the quarter note length as specified in the MIDI file; this *normalised average quantisation deviation* gives us a reasonable estimate of the quantisation error.

#### 4.1.3 Selection of relevant MIDI files

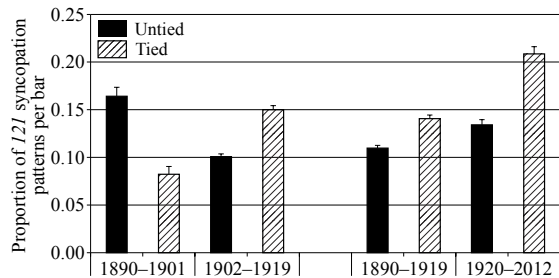
Since we want to test hypotheses about the occurrence of syncopation patterns in ragtime compositions stemming from specific time periods, we need to match the metadata in the compendium to the MIDI files in the RAG-collection. This is done by comparing the file name with the title field in the compendium. To make the file name comparison more robust, we removed all spaces, commas, parentheses etc. before doing a case-insensitive comparison. Moreover, not all MIDI files are suitable for analysis: we require that a MIDI file has only one metre, which should be  $\frac{2}{2}$ ,  $\frac{2}{4}$ , or  $\frac{4}{4}$ , and starting exactly at tick 0. Also, we found that pieces after 1920 have on average a higher normalised average quantisation deviation than pieces before 1920. To eliminate this bias, we removed all pieces that have a normalised average quantisation error above 2% from the corpus. For many entries in the compendium (a single entry stands for a unique ragtime composition) there exist multiple MIDI files in the collection. In order to not bias our study by using different MIDI renditions of the same ragtime composition, we select only one MIDI rendition per composition, namely the file with smallest average quantisation deviation.

We divided the data in two periods: *the ragtime era* period (1890–1919, 1806 pieces) and the modern period (1920–2012, 659 pieces). To be able to investigate changes in syncopation patterns in different time periods, we split pieces from the ragtime era into the pieces up to 1901 (253 pieces) and from 1902 onwards (1553 pieces).

## 4.2 Testing of syncopation hypothesis

For testing Berlin’s hypothesis on syncopations in the ragtime era we calculate the proportion of tied and untied *121* syncopation patterns in the period 1894–1901 (253 pieces) and the period 1902–1919 (1553 pieces). All pitch information of the ragtime melodies is ignored, and we only examine the onsets of the melodies. The melodies are segmented in two bar segments that overlap exactly one bar.<sup>4</sup> Furthermore, based on the time signature, the patterns are scaled appropriately to match the predominant pulse. As a consequence, in  $\frac{2}{2}$  and  $\frac{4}{4}$  metres the patterns

<sup>4</sup> This is necessary since the tied pattern in Fig. 1(e) crosses a barline.



**Figure 2.** On the left side, the proportion of untied and tied *121* syncopation patterns per bar averaged over all songs before and after 1902 is displayed. On the right side we show the distribution of tied and untied syncopation patterns for the ragtime era and for modern ragtimes. The error bars denote the standard error of the mean.

in Figure 1 (a)–(e) are matched at the quarter note level. We match all tied and untied patterns for all songs written before and after 1902, and average the results. The experiment was implemented using the functional programming language Haskell, and the code is available to the research community on request.

The results displayed on the left side of Figure 2 show that the number of tied patterns has almost doubled in the period 1902–1919, compared to the period 1890–1901. In the same period, the number of untied patterns has decreased with approximately one third. We examined if these differences were statistically significant by performing two Wilcoxon rank-sum tests, and we can confirm that the difference in mean before and after 1902 are statistically significant for the untied pattern,  $p < 0.0001$ , as well as for the tied pattern,  $p < 0.0001$ .<sup>5</sup>

Examining the difference in tied and untied patterns in the entire ragtime era (1890–1919, 1806 pieces) reveals that the differences are less pronounced (see the right side of Figure 2). The average *121* proportion per bar is 0.11 for untied and 0.14 for tied syncopation. We confirm that this difference is still statistical significant at the  $\alpha = 0.01$  level with a Wilcoxon signed-rank test,  $Z = -5.57, p < 0.0001$ . If we examine the difference between the proportions of tied and untied patterns after the ragtime era (1920–2012, 659 pieces), we observe that the usage of syncopation patterns increases compared to the ragtime era. Hence, we cannot confirm the hypothesis that a consolidation of the ragtime style within the ragtime era led to the same amount of syncopation patterns after 1919 (except for the fact that tied syncopation occur more often than untied syncopation); neither can we observe a decreased use of these patterns, which would be in accordance with Berlin’s hypothesis that the increased use of dotted rhythms in late ragtimes reduced the need for syncopation. Especially the proportion of tied syncopation increases significantly, and the difference between tied and untied syncopation is highly statistical significant,  $Z = -8.62, p < 0.0001$ . The increased use of these synco-

<sup>5</sup> One might think that the large difference in sample size has affected the significance tests. We reassure the reader that the differences remain statistically significant if we take a random sample of 253 songs after 1902 for tied,  $p < 0.0001$ , and untied,  $p < 0.0001$ , patterns.

pation patterns in the modern era indicates that our second hypothesis on the habituation to syncopation by the general musical audience in the 20th century allowing an even stronger use of syncopation in ragtimes, should indeed not be rejected on the basis of the RAG-collection. However, since the RAG-collection is much more complete for the ragtime era than for the modern era, it is difficult to predict whether this result is representative for all existing modern ragtime pieces (including those for which no MIDI file exists yet).

## 5. CONCLUSION AND FUTURE WORK

In this paper we have provided an overview of open questions on ragtime as a genre, and have argued on how MIR-methods can productively contribute to answering these questions. As a first step for finding characteristic musical features of the ragtime genre, we have evaluated musical hypotheses on syncopation patterns on the RAG-collection. We introduced the RAG-collection that evolved from the shared effort of the Internet community of ragtime lovers, and introduced a number of preprocessing steps to be able to parse the ragtime files of this collection for a corpus-based study; specifically, we tailored the skyline melody extracting algorithm successfully to ragtimes.

With our computational investigation of the occurrence of the *121*-pattern we were able to confirm Berlin's hypothesis on the increased use of tied syncopations in comparison to untied syncopations in the ragtime era after 1902. Furthermore, by extending our investigation to modern ragtimes, we were able to show that these patterns remain important in ragtime also after 1920. We have found no evidence that the amount and kind of syncopation after the ragtime era stays stable, rather we observe an increased use of these patterns, especially of tied syncopations. The increase of syncopation in ragtime over time concluded from our analysis of 2465 compositions provides an interesting complementary finding to the increase of syncopation in American popular music in 1890-1939 concluded in [9] from short samples extracted from 437 pieces.

In the near future, we will quantitatively test all occurring kinds of syncopation patterns in ragtime and determine the proportion of the *121*-pattern among these patterns. This will contribute to evaluating whether the *121*-pattern can indeed be considered such a typical pattern in ragtime, providing a promising candidate for typical ragtime features for classification tasks in MIR. For modelling syncopation patterns as a characteristic musical feature of ragtime, another next step is to consider the location of these patterns within specific form sections, such as discussed in [7] in order to establish different typologies of ragtimes. Apart from rhythmic features, musicologists argue that there exist also typical harmonic patterns in ragtime, such that the study of chord progressions is a next logical step. The strengthening of the syncopation by placing a chord rather than only a melody note at the right hand part provides yet another candidate of syncopation patterns typically for ragtimes that we are going to evaluate.

For distinguishing syncopation patterns typical for rag-

times and testing their discriminative power for automatically classifying ragtimes, we plan to extend our database with jazz, pop and rock pieces, which constitute related genres for which syncopation has been claimed being relevant [9, 13]. In a study of syncopation in rock, Temperley argues that a "full understanding of rock requires consideration of social and cultural aspects". This certainly applies also to ragtime music, and by employing content-based MIR-methods for genre classification, we plan to contribute to the debate on how far a musical genre is both a sociological construct and is rooted in musical characteristics that can be extracted from the musical content.

**Acknowledgments.** We acknowledge the efforts of many ragtime enthusiasts that have contributed to the RAG-collection and the accompanying Ragtime Compendium. Specifically, we thank Michael Mathew, who has kindly provided us with the collection and the compendium. We thank F. Wiering and M.E. Rodríguez-López for providing valuable comments on an earlier draft on this text. A. Volk and W.B. de Haas are supported by the Netherlands Organization for Scientific Research, through the NWO-VIDI-grant 276-35-001 to Anja Volk.

## 6. REFERENCES

- [1] J. J. Aucoutourier & E. Bigand: "Mel Cepstrum & Ann Ova: The difficult dialog between MIR and Music cognition." *ISMIR-Proceedings*, pp. 397–402, 2012.
- [2] E. A. Berlin: "Ragtime. A Musical and Cultural History." University of California Press, 1980.
- [3] E. A. Berlin: Review of "That American Rag: The Story of Ragtime from Coast to Coast", by David A. Jasen and Gene Jones. *American Music*, Vol. 19, No. 4, pp. 474–476, 2001.
- [4] S. DeVeaux: Review of "Ragtime: Its History, Composers, and Music" by John Edward Hasse; "Reflections and Research on Ragtime" by Edward A. Berlin. *Ethnomusicology*, Vol. 32, No. 2, pp. 142–145, 1988.
- [5] P. Dickinson: "The Achievement of Ragtime: An Introductory Study with Some Implications for British Research in Popular Music". *Proceedings of the Royal Musical Association*, Vol. 105, No. 2, pp. 63–76, 1978-1979.
- [6] F. Gillis: "Hot Rhythm in Piano Ragtime". *Music in the Americas*, pp. 91–104, 1967.
- [7] I. Harer: *Ragtime. Versuch einer Typologie*. Verlag Hans Schneider, Tutzing, 1989.
- [8] H. Honing & W.B. de Haas: "Swing Once More: Relating Timing and Tempo in Expert Jazz Drumming", *Music Perception*, Vol. 25, No. 5, pp. 471–476, 2008
- [9] D. Huron & A. Ommen: "An Empirical Study of Syncopation in American Popular Music, 1890-1939", *Music Theory Spectrum*, Vol. 28, No. 2, pp. 211–231, 2006.
- [10] M. Mathew: "A Ragtime Compendium" <http://ragtimecompendium.tripod.com/>
- [11] M. Mauch & S. Dixon: "A corpus-based study of rhythm patterns", *ISMIR-Proceedings*, pp. 163–168, 2012.
- [12] L.B. Meyer: *Style and Music: Theory, History, and Ideology*, University of Chicago Press, 1989.
- [13] D. Temperley: "Syncopation in Rock: A Perceptual Perspective", *Popular Music*, pp. 19–40, 1999.
- [14] A. Volk, W.B. de Haas, & P. van Kranenburg: "Towards Modelling Variation in Music as a Foundation for Similarity", *Proceedings of the 12th ICMPC*, pp. 1085-1094, 2012.
- [15] A. Volk, & A. Honing: "Mathematical and computational approaches to music: challenges in an interdisciplinary enterprise" *Journal of Mathematics and Music*, Vol. 6, No. 2, pp. 73–81, 2012.
- [16] A.L. Uitdenbogerd & J. Zobel: "Manipulation of music for melody matching." *In Proceedings of the ACM Multimedia Conference*, pp. 235-240, 1998.