STRENGTHENING INTERDISCIPLINARITY IN MIR: FOUR EXAMPLES OF USING MIR TOOLS FOR MUSICOLOGY

A.K. Honingh[1], J.A. Burgoyne[1,2], P. van Kranenburg[3] and A. Volk[4]

[1]University of Amsterdam, the Netherlands
[2]Netherlands Institute for sound and vision, the Netherlands
[3]Meertens Institute, the Netherlands
[4]Utrecht University, the Netherlands

ABSTRACT

Music Information Retrieval (MIR) is a fundamentally interdisciplinary field. Nonetheless, a number of presentations at previous ISMIR conferences have noted that there are some fields to which MIR seems to have a natural connection but with which there have been relatively fewer collaborations. Musicology is one of the most commonly cited fields where there are good opportunities for more interaction with MIR, and this paper presents four examples of using fundamental MIR concepts and software tools (the MIR and MIDI toolboxes) to start such collaborations. The four examples cover a wide range of musicological periods, from religious chant to 20th-century pop music, and also a wide range of MIR techniques, from concepts based on symbolic data to audio-only methods that avoid the concept of a musical score altogether. We hope that other researchers may extend or adapt our examples to answer their own musicological questions and foster their own collaborations between MIR and musicology.

1. INTRODUCTION

The field of Music Information Retrieval (MIR) has been envisioned as a multidisciplinary research area, with musicology as one of its contributing disciplines [8]. Musicology has a clear importance for MIR since for automatically processing music, it is important to understand both musical contents (in terms of music theory) and the function of music in/as culture. Nevertheless, true interdisciplinarity is still often a problematic issue [1,6,27,28]. One of the problems is, how to interpret features developed in MIR in musically meaningful ways [1], such that MIR research might contribute to musicological research. Another problem is that often it remains unclear what has been learned with machine-learning approaches [18]. While Aucouturier and Bigand [1] discuss why audio features developed in MIR are not yet in the state of contributing meaningfully to music cognition research, in this article, we will show that there are musicological questions that *can* be addressed successfully with MIR tools. Hereby we will focus on the MIR toolbox [17] and MIDI toolbox [9] that have been created for a Matlab environment.

The MIR toolbox offers an integrated set of functions dedicated to the extraction of musical features such as tonality, rhythm and structures from audio files. The MIDI toolbox offers a set of functions for the purpose of analyzing and visualizing MIDI files, and contains cognitively inspired techniques for analysis within topics such as similarity, key-finding and segmentation.

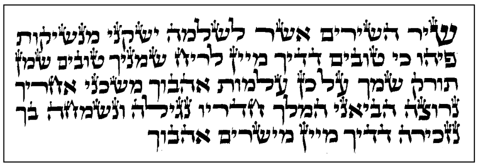
Using computational methods in musicology can bring great benefits. First of all, having digital data and computational techniques would enable large-scale analyses to test musicological hypotheses. For example, the well-known hypothesis that melodies tend to exhibit an arch shape was tested and confirmed on 36000 melodic contours using computational analysis [14]. Furthermore, computational methods would make it possible to formulate implicit musicological knowledge in an explicit, testable way. For example, the implicit knowledge about accents in music, and knowing how to dance of clap along with the music, has been made explicit and tested using a computational model that calculates the locations of the accents in a piece of music [26].

For MIR researchers, on the other hand, it is important to realize that cross-disciplinary collaboration is increasingly important to successful research projects, which challenges MIR to find a balance between features that are powerful but also make sense to collaborators who may not be experts in machine learning or audio signal processing. The same applies to our choice of learning algorithms: A fellow scholar who misses the intuition behind the learning algorithm will be unable to contribute meaningfully to understanding its results and how they might be relevant either commercially or to other academic disciplines.

In the following sections we present four examples of research questions relevant to musicology that can be addressed with existing tools, showing that in the interdisciplinary discourse it is important to first of all find the appropriate research questions we can work on together in this field.

2. TRANSCRIPTION OF CHANT RECORDINGS

Most religions have a tradition of reading or reciting their sacred texts. One way to reinforce the special status and importance of these texts is to add a musical—or semi-musical—layer to the performance of the reading. Many of these reciting traditions have roots that go deep into his-



שיר השירים אשר לשלמה ישקני מנשיקות
פיהו כי טובים דדיך מיין לריח שמניך טובים שמן
תורק שמך על כן עלמות אהבוך משכני אחריך
נרוצה הביאני המלך חדריו נגילה ונשמחה בך
נזכירה דדיך מיין מישרים אהבוך

Figure 1: Incipit of the book *Song of songs*.

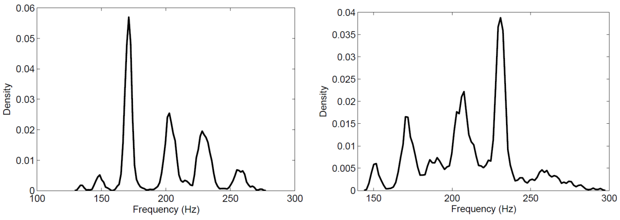
tory. As the transmission of exact melodies, modes and pitches happens in most cases orally, the reconstruction of historic renditions of these texts is often impossible. The present plethora of chant styles is the result of a very complicated history. This applies to Christian chant, Jewish cantillation, Qur'an recitation and many others. [4].

In this case study, we focus on two readings of the same passage from the Hebrew Bible from two different traditions, namely an Ashkenazy rendition from Hungary, and a Sephardic one from Morocco. Jewish chant, or *cantillation*, is an interesting type of chant to study for various reasons. The notation of the melodic patterns was standardized by Masorite rabbis during the sixth to the tenth centuries based on existing oral traditions. This notation consists of thirty graphic signs (*te'amim*) that were added to the Hebrew text to indicate syntactic relations between the words. The signs also serve to indicate melodic shape or movement for the reading. Although the notation was standardized, the rendition of this notation was not. The exact rendition of the signs is determined by various factors, such as the liturgical context, regional traditions, and improvisatory elements incorporated by a given "reader" [25].

During the twentieth century, many audio recordings have been made of Jewish readings all over the world. One way to study the diversity of chant traditions is by analyzing those recordings. We now focus on two readings of the beginning of the book *Shir hashirim* (Song of songs), one from Morocco and one from Hungary. The written text is is depicted in Figure 1. This notation does not include the *te'amim*. The reader is supposed to memorize these.

A basic question when studying the various renditions of this text is what pitches the readers use and what the relative importance and function of those pitches is. For these kinds of questions, MIR offers important tools. The MIR-Toolbox contains various functions to study the pitches in an audio fragment.

The toolchain we use is the following. First, we detect the pitch contour by performing F0-detection in sequentially overlapping time windows. The MIR-Toobox provides the function `mirpitch` for that. We apply a median filter to correct for isolated outliers, and we set the maximum frequency to 400Hz since we know there is a single male voice singing. Next, we construct a pitch density estimation with MATLAB's `ksdensity` command. The respective diagrams for the Hungarian and Moroccan readings of *Shir hashirim* are depicted in Figure 2. These diagrams reveal the tone scale the reciters were using as averaged over the entire reading, which in both cases lasts for about three minutes.



(a) (b)

Figure 2: Pitch densities of the Hungarian reading (a) and the Moroccan reading (a) of the beginning of the book *Shir hashirim*.

These density diagrams leads to important hypotheses about the pitch content of the readings. Both clearly show one most frequently used pitch, which can be regarded a reciting tone. In the Hungarian tradition, most of the rest of the melody is above the reciting tone. The two lower pitches are mainly sung in cadences. In the Moroccan rendition, on the contrary, most of the rest of the melody unfolds below the most frequent pitch.

In the Hungarian rendition we find in total six distinct peaks at 133.7, 148.6, 171.1, 202.6, 228.1, and 225.0 Hz, respectively. These values were obtained by applying MATLAB's `findpeaks` command. The corresponding tones are slightly lower than d♭, e♭, f, a♭, b♭, and c^1 in modern, equally tempered tuning (a^1=440Hz). The 'gap' of a minor third between the third and fourth pitches is an interesting feature, since it is reminiscent of pentatonic modes. The highest of the six pitches, approximately c^1, does not function as leading tone as would be the case in a tonal d♭-major context, since the d♭^{1} is completely absent in the signal.

In the Moroccan recording, the pitch density diagram shows presence of more pitches between the scale tones. This relates to a less precise style of performance employing more glissandi. Furthermore, there seem to be two options for the intonation of the third tone of the 'scale' throughout the reading, since there are two very close peaks in the density diagram. However, in general, we have to be careful about drawing conclusions from the exact locations of the peaks in the density diagrams, since these densities are computed over the entire recording. There might be different ways of intoning certain scale tones or the absolute pitch might gradually change in the course of the reading, which would 'blur' the pitch density diagram.

As we see, a summary of the pitch content of a chant recording offers many important clues for understanding the particular style of performance. Because the time information is entirely lost, a next step would be to study the succession of pitches in both chants. This would reveal functional relationships between the pitches. One such study has been done previously [16] were we showed that the rendition of the individual *te'amim* is more stable within the Hungarian (Askenazy) rendition than in the Moroccan (Sephardic) rendition.

3. ENTROPY CHANGE IN HAYDN'S STRING QUARTETS

It has often been said that Haydn's string quartets have changed in style since his Opus 33. Haydn himself has written letters about this before publishing his Opus 33 string quartets. Others have assessed his string quartets and have written about the nature of the change of composing, and about the question whether there really was a substantial change at this Opus number and not already before, at Opus 20 [11, 22]. About the nature of the change, most authors agree that there has been a change in the ratio between melody and accompaniment, in the sense that in the new style, the melody is divided more evenly over the different voices than before, when the melody was primarily to be found in the first violin [11, 13, 21, 22]. The question that remains is: when did this compositional change take place?

3.1 MIDI toolbox

We have studied this question using a corpus consisting of 30 complete stringquartets by Haydn in MIDI format, which we downloaded from `http://www.kunstderfuge.com/haydn.htm`. We have divided the string quartets, each consisting of four parts, into three groups: 1) string quartets from before Opus 20, 2) Opus 20 string quartets, and 3) Opus 33 string quartets. We have used the MIDI-toolbox [9] to study the music, and to answer the question raised above.

3.2 Cross-entropy

Entropy is a measure of randomness in a probability distribution, which has been applied to music in a number of ways [7, 15]. It has been suggested before that a melody contains a higher degree of unexpectedness than its accompaniment and would therefore have a higher entropy [7]. Using this could help us to find out how the melody is divided over the several voices of Haydn's string quartets.

If the entropy of distributions should be compared to each other, we can use the cross-entropy. Entropy and cross-entropy are to be computed over a certain distribution. Since our focus is on melody, we are concerned with the interval distribution and duration distribution. We calculated the cross-entropy using the interval distribution (function *ivdist1* from the MIDI toolbox) and duration distribution (*durdist1*) of each voice in combination with each other voice.

The closer the cross entropy is to the (normal) entropy, the more the voices are alike. We have calculated the difference between cross-entropy and entropy, also called Kullback-Leibler divergence or relative entropy, for all combinations of voices. The research question we want to address is when the composition style changed such that melody and accompaniment would be more evenly divided over the different voices than before. We can now translate this research question in terms of (cross) entropy. We are interested to find where the difference between cross-entropy and entropy, for the first violin in combination with each other voice, does decrease significantly. For completeness we will calculate the difference in relative entropy for all combinations of voices, for the three groups mentioned in the previous section. However, given the fact that, before the compositional change, the melody was foremostly present in the first violin, we expect to find more changes in the degree of similarity between the first violin and the other voices, than in other combinations of voices.

3.3 Results

We found no significant changes in the distributional similarity between voices in the Opus 20 set and the Opus 33 set. We did find some significant changes (decreasing values) in the distributional similarity between voices in the pre-Opus 20 set and the Opus 20 set. The differences of cross-entropy and entropy among pre-Opus 20 were significantly higher from the ones in Opus 20 for

1. violin 1 and violin 2 using their interval distribution ($p < 0.05$)
2. violin 1 and viola using their duration distribution ($p < 0.02$)
3. violin 1 and cello using their duration distribution ($p < 0.0001$)

This means that in the stringquartets of Opus 20 and later, 1) the score of violin 1 and 2 are more similar with respect to the entropy of their interval distribution than before, 2) the score of violin 1 and viola are more similar with respect to the entropy of their duration distribution than before, and 3) the score of violin 1 and cello are more similar with respect to the entropy of their duration distribution than before.

To summarize, we found that some distributional differences between violin 1 and the other voices are smaller in Opus 20 than before, suggesting that violin 1 is more alike the other voices in Opus 20 than before. We can interpret this as an indication that from Opus 20 onwards, the scores of violin 2, viola and cello include more elements of the melody. These results are supporting the idea that the above mentioned compositional change in Haydn's string quartets did not appear in Opus 33, but did already appear in Opus 20.

Haydn's string quartets Opus 3 are nowadays attributed to Romanus Hoffstetter [24]. According to this information we should therefore exclude Opus 3 from our pre-Opus 20 corpus, in order to focus solely on the compositional intentions of Haydn himself. If we remove Opus 3 from our pre-Opus 20 corpus set and run the experiment again, the significance of 1) disappears but another significant difference appears, namely:

4. violin 1 and cello using their interval distribution ($p < 0.05$),

which supports the same (summarizing) interpretation as before.

We have shown that, using the concepts of entropy, cross-entropy, interval- and duration-distributions from the MIDI-toolbox, we have been able to formalize part of the hypotheses that existed in the literature on the question about compositional style change in Haydn's string quartets. In this way, computational modeling can help to strengthen a presumption and can support an analysis.

4. INVESTIGATING MUSICAL PERFORMANCE STRATEGIES IN CASES OF METRIC AMBIGUITIES

4.1 Musicological problem statement

What are suitable performance strategies for a piece if metric ambiguities occur? There exist numerous musical examples where the notes of a composition imply a different metric structure than the notated bar lines, creating interesting metric discrepancies, such as, for instance, in the works by Schubert, Schumann, Haydn, Brahms, Stravinsky and Webern [10, 12]. Musicologists argue that "somehow" the performer has to cope with these ambiguities, such that he/she has to convey these ambiguities to the audience [3, 10]. For instance, Epstein reports that musicians sometimes ask whether the notated bar line is of any importance for the performance in cases where the actual music is not in sync with the notated metric grid, arguing that the notation is not relevant [10, p. 13]. However, Epstein states that the notated metric grid is in these cases of special importance and that the performer needs to convey this metric discrepancy to the listeners.

However, it remains open *what* exactly musicians should do to convey these metric ambiguities to the audience. Would the performer need to emphasize with extra loudness accents on the first downbeat of the bar where the notated downbeat is located - to make the ambiguity audible? We will investigate this question in a well-known example of metric ambiguity, namely Schubert's Moment Musical op. 94, No. 4.

4.2 Metric ambiguities in Schubert's Moment Musical

The middle part of Schubert's Moment Musical Opus 94, No. 4 contains metric ambiguities in the following sense. Both melody and rhythm of the middle section in D flat Major seem to suggest that the actual downbeat is located on the quarter note at the third occurring onset in each bar (the first four bars are shown in Figure 3). This note forms a syncopation, since it is placed on a metrically weak position, while the following strong second beat position of the bar has no note onset.

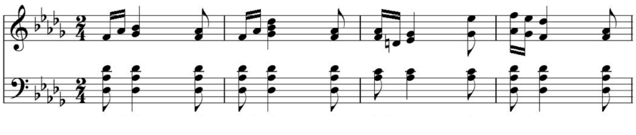
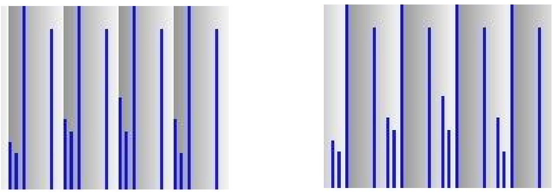



Figure 3: Bars 62-65 of Schubert's Moment Musical Op. 94 No. 4 (Beginning of the middle part)

In [26], it was shown that *Inner Metric Analysis*, a computational model for the *inner* metric structure of a musical piece (as opposed to the *outer* metric structure given by the notated time signature and bar lines) reveals that great metric accents for this passage are located not on the first beat of the bars, but on the third occurring onset in all bars. Figure 4 displays the weight profile generated from a symbolic encoding of the piece (for the model underlying the calculation of the weights see [26]). The higher the line, the greater the corresponding weight of the onset; the grey lines in the background indicate where the bar lines are located. The weight profile in Figure 4b shows the same weight profile as in 4a but with shifted bar lines towards the greatest metric weights, revealing that a typical accent pattern corresponding to a 2/4 is exhibited in the weight profiles, yet it is shifted due to the syncopation.



(a) Metric weight profile:

(b) Metric weight profile shifted:

Figure 4: Metric weight profiles of bars 62-65 of Schubert's Moment Musical. Left: Bar lines in the background as notated. Right: bar lines shifted towards the greatest metric weight.

Following Epstein's argument that the performer has to convey the metric ambiguity to the audience, an interesting performance hypothesis for this piece is then whether a performer intentionally emphasizes the notated first beat of the bars. Otherwise a listener might simply perceive the quarter note (the third onset) with the great metric weight as the downbeat and hence not recognize the metric discrepancy at all.

4.3 Analysis of a musical performance of Schubert's Moment Musical with the MIR-Toolbox

As a case study, we analyze the performance of Schubert's middle part by Maria Joao Peres [20], specifically the first phrase (eight bars) that contains these syncopations throughout. We test whether the performer stresses the notated first beat of the bar in the performance parameter *loudness* using the `mironsets` function from the MIR-toolbox that returns the onsets in the audio file along with their amplitude value. [1] Figure 5 reveals that no performance emphasis in loudness is put on the notated downbeat according to the bar lines in general, as the low amplitude values demonstrate, rather the performer seems to emphasize in many bars the syncopation (high amplitude values of the corresponding onsets on the syncopated note). Since in the

---

[1] Although the relation between amplitude and perceived loudness is not one-to-one [19], for the frequency range that is considered here, the relative loudness can be represented by the relative amplitude.

very first bar of this excerpt not all onsets are captured by `mironsets`, we discuss only bars 2-8.

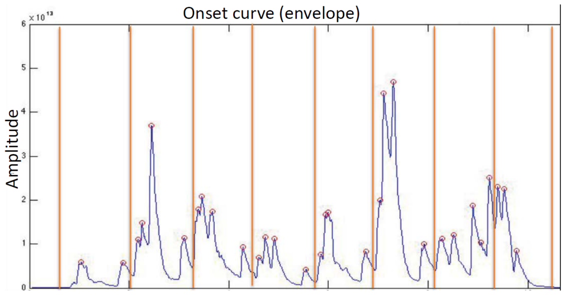


Figure 5: Excerpt for detected onsets (x-axis) and their amplitude (y-axis) for the first 8 bars using the `mironsets` function from the MIR-toolbox. Vertical lines indicate the locations of the notated bar lines.

Table 1 displays the relation of amplitude between the first note within each bar (*1st onset*) and the note located at the third eighth note position (*3rd onset*) in terms of their ranks regarding the amplitudes of all notes in the bar. If the performer stresses *1st onset* in order to convey the notated downbeat, then this onset should have the highest amplitude (rank 1) in the corresponding bar, or at least it should be louder than *3rd onset*. None of this is the case: *3rd onset* is in most bars louder than *1st onset*, showing that the performer puts more emphasis on the syncopation. Hence, our hypothesis is not confirmed: the performer does not put extra accentuations on the notated downbeat to convey the beginning of the bar, except for the last bar, the ending of the phrase. This might indicate the following: while a listener might misinterpret during the first 7 bars of the phrase that the downbeat is on *3rd onset*, this interpretation is questioned at the end of the phrase by the emphasis of *1st onset*, leading to a possible re-interpretation what has been heard so far. This might indeed be a means of expressing the ambiguity of this passage, thus delivering a new hypothesis that would be worth testing in a more extensive study than this case study.

Bar	2	3	4	5	6	7	8
Rank 1st onset	4th	2th	3rd	4th	3rd	4th	1st
Rank 3rd onset	1st	3rd	2nd	1st	1st	3rd	3rd

Table 1: Rank of the 1st and 3rd onsets per bar according to amplitude.

5. TRENDS IN THE BILLBOARD HOT 100

Musicologists are often interested in understanding how particular musical styles evolved throughout important historical periods. Trends in popular music have received particular attention since Joan Serrà and colleagues published a study suggesting that over the past half century, the range of timbres used in popular music has become more restricted and the average volume has increased [23]. This study was based on the Million-Song Dataset, which was first announced at this conference two years ago [2]. The sampling strategy underlying the Million Song Dataset, however, is more suitable for studying user behaviour or contemporary tastes than it is for history. [2] We decided to examine the McGill *Billboard* data set, which was sampled to facilitate historical inquiry, to see how it compares [5].

5.1 Method

Using the MIR Toolbox `'Frame'` option, we divided each of the 683 songs that have been released from the McGill *Billboard* set so far into Hamming-windowed frames of 200 ms, overlapping by 100 ms, and chose 500 of these frames at random from each song. On each of these frames, we then computed nine features that seemed to be most understandable from a musicological perspective: `mirrms`, the root mean square amplitude (volume) of the signal; `mircentroid`, the 'average' frequency of the spectrum; `mirspread`, the standard deviation of the frequency spectrum; `mirskewness`, how much the spectrum is biased toward high or low frequencies; `mirkurtosis`, the 'peakedness' of the spectrum; `mirflatness`, how noisy the frame is (as opposed to pitched); `mirbrightness`, the relative proportion of high frequencies in the spectrum; `mirroughness`, psychoacoustic roughness; and finally `mirregularity`, short-term variation in timbre. For each of these features, we fit random-intercept models:

$$f_{ij} = (\beta_0 + \zeta_i) + \beta_1 t_i + \epsilon_{ij} \qquad (1)$$

where f_{ij} is the feature value for frame j of song i, β_0 is a fixed intercept, ζ_i is deviation of song i from the overall trend on average, β_1 is the slope of the trend, t_i is time (the date of the chart from which song i was taken), and ϵ_{ij} is the deviation of frame j from the average for song i.

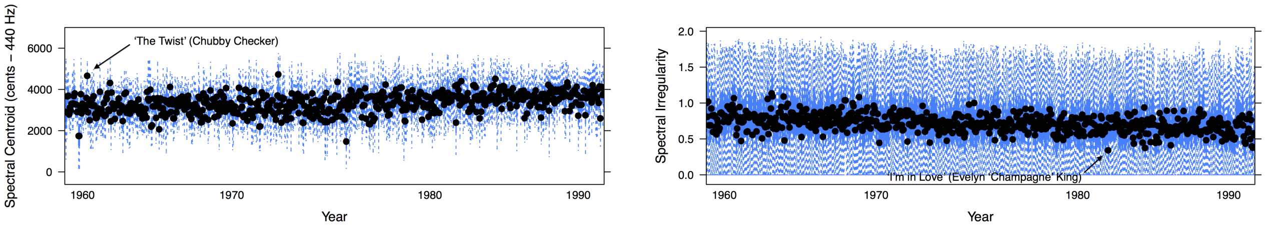
5.2 Results and Discussion

All features other than `mirrms` show highly significant trends ($p < .001$) over the period covered in the data (1958 to 1991), but the two strongest statistical relationships are for `mircentroid`, which increased at a rate of 197 cents per decade, and `mirregularity`, which decreased at a rate of 0.05 units per decade. The next two most significant trends are for `mirbrightness` and `mirflatness` (also $p < .001$), but these two features are so correlated with `mircentroid` ($r = 0.77$ and 0.88, $p < .001$) that they seem to be measuring the same phenomenon.

The most surprising result was that, contrary to Serrà et al., we found no significant relationship between `mirrms` and time. This finding depends critically on our choice of the random-intercepts model. The random-intercepts model is more conservative than classical linear regression, but we believe it is also more accurate here because it accounts separately for differences in the overall volume between different songs – which are the units that were actually sampled to create the Million-Song and *Billboard* data sets – and differences in volume within individual songs.

---

[2] `http://labrosa.ee.columbia.edu/millionsong/pages/how-did-you-choose-million-tracks`



(a) Trend in spectral centroid

(b) Trend in spectral irregularity

Figure 6: Two spectral trends in the *Billboard* Hot 100. The background of each figure includes solid lines between the first and third quartiles of the feature for each song and a dashed line to the minima and maxima. The foreground includes a point for the mean value of the feature for each song. Centroid has gradually increased while irregularity has decreased.

Nonetheless, the effect of loudness in the Million-Song set was so strong that the question of increasing volume deserves further study.

Spectral centroid, brightness, and flatness all seem to relate to how bright a timbre will be perceived to be, and their high inter-correlations support that hypothesis. Centroid is perhaps the most musically intuitive among the three as it represents a sort of average pitch, albeit biased upward due to the presence upper partials in any musically plausible spectrum. Figure 6a shows that centroid is a tight feature, with relatively little variation within individual songs. But why should it be increasing? The songs that deviates most from the trend offers a clue: Chubby Checker's 'The Twist' (1958) is the furthest above the trend of any song in the data set, and it is also the song that was deemed the most popular of all time on *Billboard*'s 50th-anniversary Hot 100 chart. We plan to investigate whether there is a more systematic link between centroid and popularity; if so, it makes sense that centroid would increase over time.

Figure 6b shows that spectral irregularity can sometimes take extreme values, which is musically plausible: A song that maintained a single, highly consistent timbre from start to finish would normally have little musical interest. The solid background lines showing the inter-quartile ranges, however, are tighter around the means and show that most of the time, pop songs maintain a fairly constant amount of local variability in timbre. We found that this amount of local variability has been falling, which is consistent with Serrà et al.'s findings. The song furthest from the trend is Evelyn 'Champagne' King's 'I'm in Love' (1981), which is indeed prototypical of a somewhat bland and synthetic sound that was characteristic of American pop in the 1980s. The connection to popularity is less obvious here, but given our findings on centroid, we plan to investigate whether lower irregularity also contributed to popularity, which would be a natural explanation for the trend.

6. CONCLUDING REMARKS

We have argued that the collaboration between the fields of MIR and musicology could be strengthened. A first attempt to do this is by focusing on MIR features and methods that can be more easily interpreted in musicological terms. We have presented four case studies that all have the potential for more extensive and data-rich investigations as follow-up studies. They provide a spectrum of research questions from different areas in musicology that inform MIR-researchers on the type of questions musicologists are interested in, and hopefully inspire researchers for setting up their own collaborations.

7. ACKNOWLEDGMENTS

The authors wish to thank Leendert van der Miesen for the idea of looking into the topic of style change in Haydn's string quartets. AKH, AV and JAB are supported by the Netherlands Organization for Scientific research. AKH is supported by an NWO-VENI grant (639.021.126), AV by an NWO-VIDI grant (276.35.001), and JAB by an NWO-CATCH grant (640.005.004). PvK is supported by the Royal Netherlands Academy of Arts and Sciences.

8. REFERENCES

[1] J. J. Aucouturier and E. Bigand: Mel Cepstrum & Ann Ova: The Difficult Dialog Between MIR and Music Cognition. *Proceedings of ISMIR*, pp. 397-402, 2012

[2] T. Bertin-Mahieux, D.P. Ellis, B. Whitman, and P. Lamere: The million song dataset. *Proceedings of ISMIR*, 2011

[3] W. Berry: *Musical Structure and Performance*, Yale University Press, New Haven and London, 1989.

[4] D.P. Biró: *Reading the Song: On the Development of Musical Syntax, Notation and Compositional Autonomy*diss., Princeton University, 2004.

[5] J. A. Burgoyne, J. Wild and I. Fujinaga: An expert ground-truth set for audio chord recognition and music analysis. *Proceedings of ISMIR* 2011.

[6] N. Cook: Towards the complete musicologist. *Proceedings of ISMIR*, 2005.

[7] B. Duane: Information Content in Melodic and Non-melodic Lines. *Proceedings of ICMPC*. 2010.

[8] J. S. Downie: Music information retrieval. *Annual review of information science and technology*, 37(1), 295-340, 2003.

[9] T. Eerola and P. Toiviainen: MIDI Toolbox: MATLAB Tools for Music Research. University of Jyväskylä: Kopijyvä, Jyväskylä, Finland, 2004.

[10] D. Epstein: Brahms und die Mechanismen der Bewegung: die Komposition der Auffuehrung. *Brahms-Studien*, Vol. 10, 1994, p. 9-21.

[11] L. Finscher: *Studien zur Geschichte des Streichquartetts: Die Entstehung des klassischen Streichquartetts. Von den Vorformen zur Grundlegung durch Joseph Haydn*. Bärenreiter, 1974.

[12] A. Fleischer (Volk): *Analytic Interpretation. Steps towards the Exploration of a Research Field exemplified on Metric Structures*, PhD thesis, Humboldt University Berlin, 2002.

[13] Grave, F., and Grave, M. G. (2006). The string quartets of Joseph Haydn. Oxford University Press.

[14] D. Huron: The melodic arch in Western folksongs. *Computing in Musicology*, 10, 3-23, 1996.

[15] L. Knopoff and W. Hutchinson: Entropy as a measure of style: The influence of sample length. *Journal of Music Theory*, 27(1), 75-97, 1983.

[16] omitted for anonymous review

[17] O. Lartillot, P. Toiviainen: A Matlab Toolbox for Musical Feature Extraction From Audio, *International Conference on Digital Audio Effects*, Bordeaux, 2007.

[18] A. Marsden: ‘What was the question?Š: Music analysis and the computer. *Modern Methods for Musicology*, 137-148, 2009.

[19] B. C. J. Moore: An introduction to the psychology of hearing (Academic, San Diego). pp. 197-204, 2003.

[20] M. J. Peres, Schubert: Sonata, D.784 / 6 Moments Musicaux, D.780 / 2 Scherzi, D.593 [Audio CD], Deutsche Grammophon, 1989.

[21] C. Rosen: *Sonata Forms*. New York: Norton, 1988.

[22] A. Sandberger: Zur Geschichte des Haydnschen Streichquartetts. *Ausgewahlte Aufsdtze zur Musikgeschichte* I. Munich, 1922.

[23] J. Serrà, Á. Corral, M. Boguñá, M. Haro, and J.L. Arcos: Measuring the evolution of contemporary western popular music. *Scientific reports* 2, 2012

[24] Tyson, A., and Landon, H. R. (1964). Who Composed Haydn’s Op 3?. The Musical Times, 506-507.

[25] E. Seroussi, et al.: “Jewish music,” *Grove Music Online*. Oxford University Press. Web. 23 Apr. 2013.

[26] A. Volk: The Study of Syncopation using Inner Metric Analysis: Linking Theoretical and Experimental Analysis of Metre in Music, *Journal of New Music Research*, Vol. 37, No. 4, pp. 259–273, 2008.

[27] A. Volk, and A. Honingh (Eds.): Special issue: Mathematical and Computational Approaches to Music: Three Methodological reflections. *Journal of Mathematics and Music*, Vol. 6, no. 2, july 2012.

[28] Volk, A., and Wiering, F: Musicology. Tutorial in ISMIR 2011.