

This is not the version for record. The fully published version of: Ben-Tal, Oded, Sturm, Bob L., Quinton, Elio, Simonnot, Josephine and Helmlinger, Aurelie (2019) Finding music in music data : a summary of the DaCaRyH Project. In: Bader, Rolf, (ed.) Computational phonogram archiving. Cham, Switzerland : Springer Nature. pp. 191-206. can be found at [https://doi.org/10.1007/978-3-030-02695-0\\_9](https://doi.org/10.1007/978-3-030-02695-0_9)

# Finding Music in Music Data: A Summary of the DaCaRyH Project

Oded Ben-Tal, Bob L. Sturm, Elio Quinton, Josephine Simonnot, Aurelie Helmlinger

**Abstract** The international research project, “Data science for the study of calypso-rhythm through history” (*DaCaRyH*), involved a collaboration between ethnomusicologists, computer scientists, and a composer. The primary aim of *DaCaRyH* was to explore how ethnomusicology could inform data science, and vice versa. Its secondary aim focused on creative applications of the results. This article summarises the results of the project, and more broadly discusses the benefits and challenges in such interdisciplinary research. It concludes with suggestions for reducing the barriers to similar work.

## 1 Introduction to the DaCaRyH project

The *DaCaRyH* project (“Data science for the study of calypso-rhythm through history”) was a 22-month long international collaboration of ethnomusicologists (Helmlinger, Spielmann), music signal processing engineers (Sturm, Quinton), music archivists (Simonnot), software developers (Parisson), and a composer (Ben-Tal). The project was jointly funded by national organisations in both France (LABEX) and the UK (Arts and Humanities Research Council), and ran from Febru-

---

Oded Ben-Tal  
Kingston University, e-mail: o.ben-tal@kingston.ac.uk

Bob L. Sturm  
Centre for Digital Music, Queen Mary University of London e-mail: b.sturm@qmul.ac.uk

Elio Quinton  
Universal Music Group, London e-mail: elio.quinton@umusic.com

Josephine Simonnot  
CREM-LESC, UMR7186, CNRS e-mail: Josephine.SIMONNOT@cnrs.fr

Aurelie Helmlinger  
CREM-LESC, UMR7186, CNRS e-mail: aurelie.helmlinger@cnrs.fr

ary 2016 to November 2017. The overarching impetus of the project was to explore how ethnomusicology can inform the development of useful and meaningful computational methods in music recording archives. This is in contrast to past approaches appearing in the literature, where (ethno)musicologists are brought a variety of computational tools and asked to use them; or where computational methods are applied to music recording collections without any involvement of (ethno)musicologists (Spielmann et al, 2017).

*DaCaRyH* was motivated by a desire of its members to familiarize themselves with the emerging field of computational ethnomusicology, and to contribute to it. *DaCaRyH* had three principal objectives:

1. *Ethnomusicological*: To enrich the domain of ethnomusicology by integrating data science and music information retrieval (MIR) methods (Schedl et al, 2014) into ethnomusicological archives and research practices;<sup>1</sup>
2. *Computational*: To enrich data science and MIR research by integrating real ethnomusicological use cases and requirements into the design of intelligent systems;
3. *Creative*: To study the concept of musical style through a comparative diachronic analysis of a music dataset, and then the creative transformation of features extracted from the same dataset into an invented, imaginary style.<sup>2</sup>

The particular music which served as the starting point of the project was the music tradition of the steelband calypso,<sup>3</sup> and steelband music as carnival arts (Guilbault, 2007; Dudley, 2002; Birth, 2008; van Koningsbruggen, 1997) and in particular, the use of rhythm across time over a period of about 50 years in the annual *Panorama competition*. This competition was created shortly after Trinidad and Tobago became independent in August 1962 as part of a policy governmental promoting masquerade, calypso traditional narrative ballads. The state-sponsored Panorama competition has since become the most prestigious steelband stage performance in Trinidad. Owing to both its prominent and prestigious position, the Panorama competition has had a prime influence on the development of steelband music (Stuempfle, 1995; Dudley, 2008).

In this article we review the results of *DaCaRyH*, and discuss lessons we learned in the process. Section 2 reviews the context of the project, and Sec. 3 discusses its groundwork. In Sec. 4, we describe results of this project in the computational ethnomusicology domain, and in Sec. 5, we summarise our creative work. We then draw some conclusions about such interdisciplinary and collaborative research.

---

<sup>1</sup> Music information retrieval encompasses computational methods for extracting, accessing and using information in collections music recordings. Examples include estimating the tempo of music in an audio recording, determining the key of a notated piece of music, and transcribing an audio music recording into notation.

<sup>2</sup> A “dataset” is a collection of data. In our case, it is a collection of audio music recordings.

<sup>3</sup> Calypso is a style of Caribbean music that originated in Trinidad and Tobago. For a more detailed description of Calypso, see for example Cowley (1998)

## 2 Context of the project

*DaCaRyH* arises from a somewhat contentious debate regarding “data science”, “big data” and the computational study of culture (Schellenberg and von Scheve, 2012; Serra et al, 2012; Huron, 2013; Fink, 2013; Wallmark, 2013; Sturm, 2014a,b; Mauch et al, 2015; Marsden, 2015). Mauch et al (2015) argues strongly for the utility of “big data tools” for the scientific study of music. In their work, they apply computational tools to analyse excerpts of American popular music and discover three “stylistic revolutions” between 1960 and 2010. This follows on the heels of other “big data” music studies. Serra et al (2012) investigated changes along three dimensions of Western popular music over 50 years and discover in their results that the “pitch sequences” have become more restricted, the “timbral palette” has become less diverse, and that the “average loudness” has increased. Another study (Schellenberg and von Scheve, 2012) analyses the statistics of tempo, mode, duration, artist gender, and recording year of Top-40 songs since 1960, and finds tempo to be decreasing along with the use of major mode. More recently, Weiß et al (2018) explored the stylistic evolution of Western classical music from 1660-1975 by comparing tonal descriptors such as intervallic content, chord transitions, and “tonal complexity.” Their findings essentially confirm broad observations of musicologists with regards to stylistic development and boundaries during the period of study, which suggests the underlying computational framework is musically meaningful.

Some of these findings produced using “data science” and “big data”, however, are not broadly persuasive. Though they arise from computational studies of a vast number of music recordings that would strain any human listener, the “objectivity” and “relevance” of the methods employed have been challenged. A talk at a national meeting of the American Musicological Society (Wallmark, 2013) and an editorial (Fink, 2013) — both by musicologists — bemoans the “science” of such studies, the aggrandisement of such findings by the media, and the lack of participation in such studies of musicologists, i.e., those who actually *study* music and culture. Also problematic are implications that what musicologists do is not a “scientific study” of music, in contrast with “big data” approaches, which themselves have inherent biases. Recent surveys (Sturm, 2014a,b) show an enormous amount of work has been produced by music data scientists but that much of it treats music superficially, and produces conclusions with questionable scientific validity. The presentation by Marsden (2015) at the 2015 workshop *Music Similarity: Concepts, Cognition and Computation* also criticises the treatment of music as documents from which to extract a diverse number of statistics of unknown relevance, instead of treating music as artefacts or records of human culture with rich contexts.

From the published literature, it seems that musicologists have so far rarely been invited to take a seat at this “big data buffet”, but this is beginning to change. The AHRC-funded project “Transforming Musicology”<sup>4</sup> piloted ways of using digital tools in musicological studies. One project it supported was the “Digital Music Lab”

---

<sup>4</sup> <http://www.transforming-musicology.org>

(AH/L01016X/1),<sup>5</sup> which sought to build tools to facilitate “big data musicology” (Weyde et al, 2014). A case study of applying “big data” to musicology is given by the project “A Big Data History of Music” (AH/L010046/1).<sup>6</sup> The US musicologist Huron (2013) argues that it is a “moral imperative” to welcome “big music data,” and for new trainees to become versed in the methods of data science and statistics. This call to action is also echoed by Wallmark (2013).

This context motivated the approach we proposed in *DaCaRyH*: instead of starting by asking what (ethno)musicological questions might be answered by specific tools and techniques of music data science, we begin from the ethnomusicological questions, and then attempt to create bespoke tools that could help address them. The resulting learnings may then feedback and inform MIR practices as well as feeding our creative strand.

### 3 Groundwork

#### 3.1 *Steelband’s Calypso: Panorama recordings*

Calypso is a polysemic term referring to a group of related styles originating from Trinidad and Tobago. In its most common sense, it is a song of social commentary, with a specific rhythmic accompaniment, and performed in the western tonal system. The calypso beat is also the polyrhythmic accompaniment, that can be played independently from the song. Historical researches show the earlier forms were responsorial, played with a polyrhythmic accompaniment, in the carnival procession (Cowley, 1998; Regis, 1999). The 19th century musical form called kalinda first played both indoor (in tents) and in the streets, evolved differently according to the contexts, that oriented organological and musical changes, within a certain stylistic homogeneity. The calypso song appeared in the late 19th century, and is the heir of the indoor style, that adopted Western instruments, and the street songs and polyrhythms later evolved to steelbands, after several organological changes: from drum to tamboo bamboo (stamping tubes polyrhythm), and from bamboos to metallic second hand instruments.

The adoption of metal allowed a dramatic musical innovation in the 1940s: the players started to differentiate several pitches on the same surface (Stuempfle, 1995; Dudley, 2008). This inclusion of pitches in the polyrhythm launched the era of “steelpans” (or “pans”) melodic idiophones made of the 55-gallon oil drums and main instrument of steelbands. A music style had led to the invention of a new musical instrument family, that became popular in the US under the name “steeldrums”. The making of steelpans gradually professionalized, and is now highly specialized: to make a pan, the maker (called tuner) sinks the top of the drum, shapes the notes, cuts the body of the drum. Then the instrument is burnt, tuned with a hammer, some-

<sup>5</sup> <http://dml.city.ac.uk>

<sup>6</sup> <http://gtr.rcuk.ac.uk/project/3510829B-EAE9-48DC-A723-8093D92CAD60>

time chromed. Tuners perform nowadays generally the “harmonic tuning”, the tuning of some overtones following the harmonic series, which had a great impact on the tone of the steelpans (Helmlinger, 2012). Although the earlier pans were made of one drum (and they still do in the “single pan bands”), the number of drums per musician can — in the fully chromatic “conventional bands” — reach up to twelve for the lowest range instrument.

Amazingly, although calypso music became a natural music style for steelbands, it not the only one: steelbands have rapidly played in various contexts, in addition to the carnival procession. They started soon to perform a large variety of tunes, from western classical music to pop and jazz (Stuempfle, 1995). But a political event, the Independence of Trinidad and Tobago (1962), happened to influence the music scene in the steelband movement: a new competition, the Panorama, was created by the authorities as soon as 1963. Scheduled right before the carnival, its rules constraint the band to play a calypso. Panorama quickly became the major event amongst steelbands, gathering hundreds of bands (all categories included), composed of up to a hundred players each. Interestingly, the competitive context happened to shape a particular style of music, belonging to the calypso family: the Panorama tune. The rules constraint the bands to perform the arrangement of a calypso. The arrangers tasks is to select a song, and arrange it into the Panorama structure, which is influenced by the sonata form. An introduction is composed, then the verse and chorus of the calypso are presented, followed by various variations including modulations, a change into minor mode, a sort of climax called jam, another exposition to the theme, and a coda.

Steelband music offers the rare possibility to observe the evolution of a music style — including a new family of instrument and their acoustical characteristic — almost from its creation. The Panorama competition is a consistent and very meaningful category in Trinidad and Tobago, where the steelpan is a true emblem: it was officially declared National Instrument in 1992 (Regis, 1999; Dudley, 2008).

### 3.2 *Session music transcriptions*

The creative strand of the project envisioned the transformation of features extracted from the recorded steelband music dataset to create a form of music different from calypso but informed by it. Fairly soon it became apparent that extracting useful features from the noisy recordings will be very difficult. Rather than delay the work on this aspect we decided to apply the same concept to another dataset as a proof of concept: What creative paths open with access to a symbolic representation of a large collection of traditional music?

We settled on using transcriptions of traditional music referred to as “session” music — traditional music typically accompanying dance from the UK, Ireland and France, but also including other traditions, e.g., Cajun. The website [thesession.org](http://thesession.org) serves as a hub for an international community of enthusiasts for this kind of music, and provides thousands of such transcriptions. Registered users post transcriptions

of tunes in a textual representation called “ABC notation”.<sup>7</sup> An example of one transcription from this collection is below:

```
T: Off To California
R: hornpipe
M: 4/4
K: Gmaj
|:GFGB AGED|GBdg e2df|gfgd edBG|ABAG E2DE|
|G2GB AGED|GBdg e2df|gfgd edBG|ABAF G4:|
|:gfeg fedf|edef edBd|gfgd edBG|ABAG EDEF|
|GFGB AGED|GBdg e2df|gfgd edBG|ABAF G4:|
```

The fields in this transcription denote the title (T:), dance (R:), meter (M:), mode (K:), and finally the notes and repeats in the tune. This kind of music is typically learned by ear, and so such a transcription mostly serves as a reminder (which is one of the original motivations that drove the development of ABC notation). The way a tune is played is contingent on the dance it accompanies and the instrument that is playing it, and may not be explicit in a transcription. For instance, the above transcription converted to staff notation appears as the following:

#### Off To California



However, when this tune is played to accompany a hornpipe the quaver rhythm is broken (offbeat quavers are shortened) and beats 2 and 4 are emphasised.

This crowd-sourced data contains more than 23,000 transcriptions of melodies, many of which are different versions of the same tune. The tune above has seven different transcriptions in the dataset,<sup>8</sup> each with slight variations, e.g., ornamentations, an explicit hornpipe rhythm, first and second endings, and so on. The size of this dataset makes it amenable to modeling with neural networks, which we describe in Sec. 5.

<sup>7</sup> <http://abcnotation.com/wiki/abc:standard:v2.1>

<sup>8</sup> <https://thesession.org/tunes/30>

### 3.3 *CREM and Telemeta*

The audio archives of the CNRS Musée de l'Homme gather commercial and field recordings of music and oral traditions from all around the world, from 1900 to the present. These archives are among the most important in Europe in terms of quality, quantity and diversity. The "Centre de Recherche en Ethnomusicologie", a research team of the National Center for Scientific Research in France (LESC, UMR 7186), manages this intangible heritage,

Most of the ethnomusicological archives are sound recordings resulting in unique challenges in preserving and managing such a collection in spite of the technological evolution. The temporal nature of audio-visual materials used in these fields raises special issues. Because they are used for research, they must be both well indexed and also easily accessible. They must also be managed in such a way as to provide their associated metadata and controlled access. In addition, it is important to be able to visualize the sound in order to navigate through the files and to annotate them precisely, which is not as simple as annotating text or image documents. At the same time, the French academic institutions require a wide access to the raw data to justify the continued support for research projects and continuing digitisation efforts. So the challenge is to manage the intellectual property rights as well as the technical issues.

The Research Center for Ethnomusicology (CREM) and the Laboratory of Musical Acoustics (LAM), have been working together since 2007 to design an innovative and collaborative tool for easy indexing of sound files. This tool, called Telemeta, a multimedia Web platform based on an open-source software designed by the Parisson Company, has been online since 2011. Now, access to the audio archives of the CNRS- Musée de l'Homme is offered with a streaming player.<sup>9</sup> More than 43,000 recordings and metadata collected throughout the world since 1900 are available on line in the database for research and experimentation in ethnomusicology. Since 2013, the aim is to enrich Telemeta platform with MIR tools used by data scientists (from IRIT, LIMSI, LaBRI, France) in order to pave the way for a semantic search engine. To be effective and in phase with the new audio technologies, analysis tools are expected to improve musical research activities on the web database. The reflection collectively engaged by engineers and researchers on the use of the sound archives database led us to set up a national project called DI-ADEMS (Description, Indexation, Access to Ethnomusicological and Sound Documents, 2013-2016). These new tools are used by the CREM staff for indexing since 2015. Now, to index and to segment the audio content, we are helped in this long and careful process with on line tools, like start recorder detection, speech/singing detection, monophonic/polyphonic parts detection, etc. With those tools and temporal annotation markers, the CREM is developing a new collaborative work between the sound engineer and the archivist to identify the content the sound files in an efficient way.

---

<sup>9</sup> <http://archives.crem-cnrs.fr>



## 4 Ethnomusicological application

One of the work threads in DaCaRyH consisted in an interdisciplinary case study combining traditional and computational methodologies to study Trinidad steelband music in a collection of recordings of the annual Panorama competition spanning over 50 years (Quinton et al (2017)). A number of facts and trends have been identified regarding Trinidad steelband music in the ethnomusicology literature,<sup>10</sup> while some other hypotheses formulated have not been addressed with traditional methodologies. As a result, we investigated these ethnomusicological research questions through the computational lens of MIR methods (Schedl et al, 2014) to facilitate the realisation of quantitative and labour intensive studies. More specifically, we focused on trends of tempo, tuning and dynamic range over a period of 50 years of the Panorama competition. We sought to study three research questions:

1. Does the tempo of winning Panorama performances tend to increase over time?
2. Do arrangers use increasingly large dynamic ranges?
3. Has there been a change in the sonic properties of steelbands?

We assembled a dataset of recordings of most of the steelband calypsos/socas<sup>11</sup> performed by the winners (first, second and third places) for each yearly Panorama competition between 1963 and 2015 (except for the year 1979, in which the competition was boycotted). Our dataset contains 93 recordings, each one being typically 8-10 mins long, totaling about 14hrs of audio. All the digital recordings of this dataset were extracted from the “CNRS - Musée de l’Homme” sound archives,<sup>12</sup> which are accessible via the Telemeta platform Fillon et al (2014),<sup>13</sup> as described in section 3.3. This musical data includes field recordings as well as published recordings, which are a mixture of digitised analog recordings and digital recordings.

The details of our work are given in Quinton et al (2017), but in summary we automatically extracted features, such as tempo, loudness or tuning frequency, from every audio recording in order to provide a quantitative measurement to shed some light on our research questions. Not all MIR descriptors extracted are equally reliable and some did not provide robust enough evidence to draw conclusions. But when these quantitative measurement were conclusive, our findings are in line with what appears in the ethnomusicological literature.

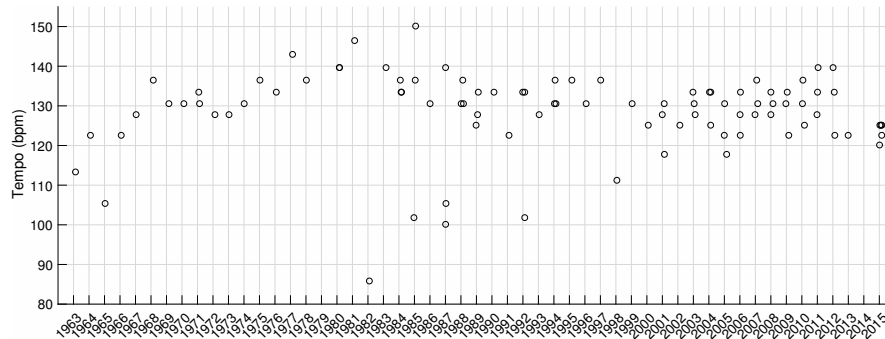
For instance, Figure 1 shows the estimated tempo of each recording in our collection in each competition year. We noticed that some of the estimates are below 110 bpm or above 140 bpm — which are unusual tempi for this music. Manual inspection reveals that most of these come from errors of the automatic tempo estimation procedure. Nevertheless, in the vast majority of cases tempo automatically extracted matches well with human hearing and may therefore be considered as reliable. On

<sup>10</sup> See for example Helmlinger (2012); Aho (1987); Dudley (2002)

<sup>11</sup> In the dictionary of the English/creole of Trinidad and Tobago edited by Lise Winer soca is defined as “a type of calypso-based music, with a fast dance beat, and party lyrics.”

<sup>12</sup> <http://archives.crem-cnrs.fr/>

<sup>13</sup> <http://telemeta.org/>



**Fig. 1** From Quinton et al (2017), the estimated tempo of each recording in our collection.

the other hand, when it comes to dynamic range and tuning, the results obtained from MIR tools on such a noisy dataset are much less reliable, and therefore do not constitute enough evidence to support a conclusion regarding our related research questions. One reason for this unreliability comes from the recording conditions: these are live events many times recorded from positions closer to the audience than the musicians. We find in several recordings that the noise from the audience has more acoustic power than the music. While the human auditory system is remarkably effective at separating such sources, this is not the case with (current) MIR tools. As a result, the numbers calculated by the automatic estimation are not musically relevant.

Music recording corpora used for ethnomusicological research typically are either ethnographic recordings or field recordings (both in our case), which are produced in very heterogeneous conditions that are mostly uncontrolled. It is often not possible to evaluate the impact of these recording conditions on the properties of the musical recordings. Our study is a good demonstration of how some MIR tools are more impacted than others by the quality of audio. Some estimates produced by MIR tools on such corpora may only have limited reliability, while some other may be reliable. As we have highlighted, results of computational analyses carried out on such a dataset should therefore be interpreted with due care when no means of assessing the reliability of feature estimation are available. In addition, our study is also a demonstration of the fact that such datasets provide a challenge for the design of MIR tools robust against uncontrolled recording conditions.

## 5 Creative application

In Sturm et al (2016) and Sturm and Ben-Tal (2017), we describe our application of machine learning to the large collection of folk melodies described in Sec. 3.2. We name this software “folk-rnn”. The method we used — recurrent neural networks (RNN, Hochreiter and Schmidhuber (1997))— take the existing context to predict

the next step in a sequence. Our first version applied this directly to the characters of the ABCnotation, the second version is based on tokens. To illustrate the difference consider the common-time meter representation:  $M: 4/4$ . In the first version it will appear as a sequence of five characters: “M”, “:”, “4”, “/”, and “4”. In the second version it is treated as a single symbol (thus “token”). Similarly a bar repeat sign “: |” would be two characters but only one token. We can see that from these two examples that the “:” character means different things in ABC notation. While an RNN might learn to take into account such differences in context, a token-based representation brings the transcriptions closer to the music these sequences encode.

After training, the resulting model is able to produce new transcriptions that share many of the characteristics of the training set. As our evaluation of this model shows (Sturm and Ben-Tal, 2017) the model encodes some compositionally useful aspects of the style including correctly counting bars, basic phrase structure, repetition and variation of patterns, and cadence points (even though the cadences themselves are not consistently good). Furthermore, experienced performers who work within this tradition did not have difficulties locating good tunes within the large collection of produced transcriptions.<sup>14</sup> The musicians performed some of these tunes together with traditional tunes in several concerts and workshops we organised as well as at a pub session in London. One of the musicians commented that the model produced interesting and stylistically appropriate patterns that he did not come across before. One interpretation of this is that machine learning can augment this musical form - allowing musicians to explore new corners of the musical space of plausible session tunes. At the same time we must also acknowledge that some would consider this a distortion of the tradition.

At the same time we explored the creative potential of this approach beyond the domain of origin. We asked two musicians who do not perform folk-music if they could find transcriptions they could play in a concert. Both John Hughes (double bass) and Torbjorn Hultmark (trumpet/soprano trombone) are experienced improvisors who regularly perform in different contexts: from orchestral works to Jazz and free improvisations. Both play instruments not typically included in session music. Yet both were able to curate and adapt transcriptions from our published collection and make them work.<sup>15</sup> This is foremost a testament to the musicality and creativity of performing musicians. But it also points to the co-creative potential of such Artificial “Intelligence” methods in the domain of music. To further explore this potential we created a web-based interface to our model. [folkrrn.org](http://folkrrn.org) allows users to generate transcriptions using our folk-rnn model. Users can generate and curate from the outputs or modify initiating parameters (such as meter, mode or opening notes) to explore the range of tunes the model is able to produce. Users will eventually be able to archive tunes they create with the tool using an online repository. There they will also be able to participate in forum discussions, browse and ‘like’ tunes other users have contributed and, hopefully, give and receive feedback about the tool and its potential for creating music. Our aim is to learn from the way others use the

<sup>14</sup> <https://highnoongmt.wordpress.com/2018/01/05/volumes-1-20-of-folk-rnn-v1-transcriptions/>

<sup>15</sup> Hultmark: <https://www.youtube.com/watch?v=4kLxvJ-rXD8> ; Hughes: <https://www.youtube.com/watch?v=GmwYtNgHW4g>

model for creative purposes to further develop this as a composition assistance tool. We will be holding a composition competition — aimed primarily at students but open to everyone — with the winning piece performed in a concert in October 2018 in London.

Both Sturm and Ben-Tal used the folk-rnn model to compose new pieces (Sturm et al, 2018). Sturm based several compositions on selected system outputs which he then arranged electronically or acoustically. He also worked interactively with the system, seeding it with an initial sequence and then curating from the resulting outputs. Ben-Tal used the system to produce pre-compositional material which was substantially edited and adapted in the composition process itself. The generation of the material was also interactive — fine tuning parameters and seeding sequences. The aim was to shift the outputs away from the core style of the model to better fit Ben-Tal’s compositional idiom. By doing that we discovered that our assumptions about what musical learning took place in the training phase were wrong. Starting the generation process from an initial sequence of tokens that deviated from the style of the training data, for example non-modal patterns, the continuation often did not contain the correct number of beats in a bar. The model was not able to apply repetition and variation on unfamiliar patterns resulting in noodling, often musically unrelated to the initial material. In other words, the ability of the model to encode relevant features is highly circumscribed and not really musical. At least not musical in human terms. The creative research strand of our project, therefore, brought to light aspects that were hidden until we started prodding our model beyond a validation of the tool with statistics.

## 6 Lessons learned

### *6.1 Differences in research practices between engineering and ethnomusicology*

Collaboration between MIR, which belongs to engineering and computer sciences, and ethnomusicology, related to both anthropology and musicology, highlights a set of differences in the scientific cultures and practices. These differences appear of course in the research methods, but also in the terminologies and practices: while MIR researchers follow engineering scientific standards which focuses on benchmarks and statistical significance, ethnomusicologists collect information through “participant observation” resulting in rich, multifaceted data most of it irreducible to numbers. Cross-disciplinary communication is therefore crucial to bridge the gap and the understanding of the potential of each others approach, and to familiarize with each others terminology, concepts, and aims. This communication is made more complex by the differences in training between the disciplines, but made possible because of the common interest — the music.

Some of the scope is shared, but each discipline is driven by concerns that may lie outside the shared scope. For instance, in the case study discussed here, both ethnomusicology and MIR are interested in timbre, loudness and tempo. But in ethnomusicological literature, it is discussed as the effects of rivalry (Dudley, 2008; Helmlinger, 2011). The rivalry, a relational characteristic observed in Trinidad and Tobago steelbands, is the aspects ethnomusicologists have focused on: in an anthropological perspective, this is the feature that allows to understand both extra-musical behaviors and musical characteristics and evolution. Tempo, timbre and loudness interest ethnomusicologist for what these features express about musical culture: they are a matter of an additional interpretation. As Rouget puts it, “la musique, c’est toujours plus que la musique” (Rouget, 1995), meaning “music is always more than just the music” and the ethnomusicological focus is wider than MIR. At the same time the scope of MIR is wider in its aims to provide answers about music in general, not just the music of a particular culture. Reconciling the differences in academic cultures requires learning to formulate (or re-formulate) research questions in ways that take into account the other fields perspectives.

## ***6.2 Difficulty in forming questions compatible with tools***

Once we narrow the scope of both disciplines to the common ground, research questions fitting in this space have to be found. A pluri-disciplinary collaboration is materialized by different aspects: research questions, methods, bibliography. Methods and bibliographic work can be just added, so they naturally mix by addition. But research questions are less easily pluri-disciplinary: the orientation of the questions depends on one’s scientific background. Goody (1977) has shown how the characteristics of the media have a very deep impact on the content of it. In our case, the tools are not just a consequence of the research questions, they also influence them. The promise of computational ethnomusicology is that with new tools come new avenues for research. Without a background in the engineering sciences, one of the difficulties has been, for the ethnomusicological part of the partnership to gain a sufficient understanding of the possibilities, as well as the limitations, offered by MIR tools to suggest ethnomusicological research questions fully adapted to the tools. Cross-readings, and extended exchanges are of course the way to overcome these hurdles: it is necessary for ethnomusicologists to be able to assimilate this new methodological arsenal in order to effectively formulate anthropological questions. Similarly, researchers working in the MIR domain would benefit from a better understanding of the wider context of (ethno)musicological knowledge to look beyond extracted features, data-sets, and ground truths.

### 6.3 *Validation of computational tools*

Computational tools designed for and applied to studying music have to be presented with a fair evaluation of their reliability or accuracy for the task in a *particular* context. In our case, manual validation shows that the tempo estimation tool is quite reliable. Most of the calculated values match what we identify from listening — the MIR tool echoes human interpretation and therefore can be considered reliable for the task. But manually double checking each MIR-produced result is little progress on estimating tempo manually to begin with. To make the tools really useful to non-MIR researchers they need to come with a “health warning”. We need a way to estimate the fitness of the tool to the items it is unleashed on. Our work on estimating tuning frequency serves as an illustration of the problem. At first, calculations from MIR tools supported our hypothesis. Further investigations, informed by an understanding of the inner working of these methods, led us to conclude that these numbers are artefacts and are not reliable.<sup>16</sup> Either any computational ethnomusicology work will depend on a MIR “technician” or the tools themselves will need to present the user with more contextual information.

## 7 Conclusion

Our project was funded by a Franco-British initiative that recognises the importance of collaboration — across disciplines and cultures — to foster research. Our experience over the last two years is that the challenges with this kind of work are often more subtle and less visible than is commonly recognised. Disciplines shape not just terminology and methodology they also influence career trajectories and pressures, academic cultures, and similar frames of reference. The two strands of work summarised in this paper illustrate two approaches to such collaborative work.

The work described in Sec. 4 fits into an interdisciplinary model where researchers in two domains bring their individual expertise to bear on a problem. The work is based on a limited dataset, where field researchers have empirical observations. Their expertise about this music allows them to posit hypothesis that can be addressed by MIR tools. At the same time, the expertise of the engineers is needed not just to apply the tools to the dataset but also to evaluate those results. Understanding the mechanisms of feature extraction is needed in order to understand how those may interact with the items in the collection.

Our work with machine learning and folk music (Sec. 5) is an example where the disciplines are much more intertwined. This outcome is the fruit of a collaboration that started almost two years before the beginning of the DaCaRyH project. Extensive discussions, most of them not tied to an immediate goal (write a paper; apply for a grant), enabled us to discover where our shared interests are and where our expertise is fruitfully complementary. The longer duration also allowed us to under-

---

<sup>16</sup> See further details in Quinton et al (2017).

stand enough of each other's methods and concerns to be able to better translate our ideas between the different academic cultures.

That suggests taking a more patient approach to allow interdisciplinary collaboration to flower. This applies to the researchers who embark on such work, bodies who fund such research, as well as academic institutions which set expectations for research "productivity". At the same time we should also expand the training in research methods in our PhD programmes. Computational methods, including both MIR and machine learning, have a role to play in future research in music. Unless we who research music and know it inside out engage with those tools others will. We need to incorporate MIR tutorials into (ethno)musicology training programmes. We need to understand the potential and limitation of these as research tools. We need to showcase instances where computation serves musical purposes and highlight cases where research is getting music wrong.

At the same time, MIR research community should consider more carefully the implicit assumptions that underlie the mechanisms they develop and include those when releasing tools. Simultaneously, MIR researchers should engage more with musicologists, ethnomusicologists, and music theorists. While it is clear that research in MIR will continue to be shaped primarily by the demands of commercial music, music departments hold stores of knowledge that is pertinent to the concerns of MIR research yet remains largely untapped.

**Acknowledgements** Florabelle Spielmann, Ghislaine Glasson Deschaumes, Andrew Thompson

## References

- Aho WR (1987) Steel band music in Trinidad and Tobago: the creation of a people's music. *Latin American Music Review/Revista de Msica Latinoamericana* 8(1):26–58
- Birth KK (2008) Bacchanalian sentiments; musical experiences and political counterpoints in Trinidad. Duke University Press
- Cowley J (1998) Carnival, canboulay and calypso: traditions in the making. Cambridge University Press
- Dudley S (2002) The Steelband's "Own Tune": Nationalism, Festivity, and Musical Strategies in Trinidad's Panorama Competition. *Black Music Research Journal* pp 13–36
- Dudley S (2008) *Music from Behind the Bridge: Steelband Spirit and Politics in Trinidad and Tobago* (illustrated edition). Oxford University Press Inc.
- Fillon T, Pellerin G, Brossier P, Simonnot J, La Dfense N (2014) An open web audio platform for ethnomusicological sound archives management and automatic analysis. In: *Workshop on Folk Music Analysis (FMA2014)*, p 36
- Fink R (2013) Big (bad) data (editorial). *Musicology Now* (online)
- Goody J (1977) *The Domestication of the Savage Mind*. Cambridge University Press

- Guilbault J (2007) *Governing sound: The cultural politics of Trinidad's carnival musics*. University of Chicago Press
- Helmlinger A (2011) La virtuosité comme arme de guerre psychologique. *Ateliers d'anthropologie* 35
- Helmlinger A (2012) Pan Jumbie. Mmoire sociale et musicale dans les steelbands (Trinidad et Tobago). *Socit d'ethnologie*
- Hochreiter S, Schmidhuber J (1997) Long Short-Term Memory. *Neural Computation* 9(8):1735–1780
- Huron D (2013) On the virtuous and the vexatious in an age of big data. *Music Perception* 31(1):4–9
- van Koningsbruggen PH (1997) *Trinidad Carnival: A Quest of National Identity*. Caribbean
- Marsden A (2015) Music similarity. In: Presentation at *Music Similarity: Concepts, Cognition and Computation*, URL <http://www.lorentzcenter.nl/lc/web/2015/669/presentations/Marsden.pptx>
- Mauch M, MacCallum RM, Levy M, Leroi AM (2015) The evolution of popular music: Usa 1960–2010. *Royal Society Open Science* 2(5), DOI 10.1098/rsos.150081
- Quinton E, Spielmann F, Sturm BL (2017) Computational ethnomusicology for exploring trends in Trinidad steelband music through history. In: Proc. CMMR
- Regis L (1999) *The political calypso: true opposition in Trinidad and Tobago, 1962–1987*. University Press of Florida
- Rouget G (1995) Ethnomusicologie d'un rituel. la représentation, ou de velasquez à francis bacon. *L'Homme* 35(133):77–97
- Schedl M, Gomez E, Urbano J (2014) Music information retrieval: Recent developments and applications. *Foundations and Trends in Information Retrieval* 8(2-3):127–261
- Schellenberg EG, von Scheve C (2012) Emotional cues in american popular music: Five decades of the top 40. *Psychology of Aesthetics, Creativity, and the Arts*, 6(3):196–203
- Serra J, Corral A, Boguna M, Haro M, Arcos JL (2012) Measuring the evolution of contemporary western popular music. *Sci Rep* 2, URL <http://dx.doi.org/10.1038/srep00521>
- Spielmann F, Helmlinger A, Simonnot J, Fillon T, Pellerin G, Sturm BL, Ben-Tal O, Quinton E (2017) Zoom arrière: L'ethnomusicologie à l'ère du Big Data. *Cahiers d'ethnomusicologie* 30:9–28
- Stuempfle S (1995) *The steelband movement: The forging of a national art in Trinidad and Tobago*. University of Pennsylvania Press
- Sturm BL (2014a) The state of the art ten years after a state of the art: Future research in music information retrieval. *J New Music Research* 43(2):147–172
- Sturm BL (2014b) A survey of evaluation in music genre recognition. In: Nürnbergger A, Stober S, Larsen B, Detyniecki M (eds) *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation*, vol LNCS 8382, pp 29–66



- Sturm BL, Ben-Tal O (2017) Taking the models back to music practice: Evaluating generative transcription models built using deep learning. *J Creative Music Systems* 2(1)
- Sturm BL, Santos JF, Ben-Tal O, Korshunova I (2016) Music transcription modelling and composition using deep learning. In: Proc. 1st Conference on Computer Simulation of Musical Creativity, Huddersfield, UK, URL <https://drive.google.com/file/d/0B1OooSxEtl0FcTBiOGdvSTBmWnc/view>
- Sturm BL, Ben-Tal O, Monaghan U, Collins N, Herremans D, Chew E, Hadjeres G, Deruty E, Pachet F (2018) Machine learning research that matters for music creation: A case study. *J New Music Research* (submitted)
- Wallmark Z (2013) Big data and musicology: New methods, new questions. Tech. rep., American Musicological Society National Meeting, Pittsburgh, PA
- Weiß C, Mauch M, Dixon S, Müller M (2018) Investigating style evolution of western classical music: A computational approach. *Musicae Scientiae*
- Weyde T, Cottrell S, Dykes J, Benetos E, Wolff D, Tidhar D, Kachkaev A, Plumbley M, Dixon S, Barthelet M, Gold N, Abdallah S, Alancar-Brayner A, Mahey M, Tovell A (2014) Big data for musicology. In: Proceedings of the 1st International Workshop on Digital Libraries for Musicology, ACM, New York, NY, USA, DLFM '14, pp 1–3, DOI 10.1145/2660168.2660187, URL <http://doi.acm.org/10.1145/2660168.2660187>