A Review of Manual and Computational Approaches for the Study of World Music Corpora

Maria Panteli, Emmanouil Benetos & Simon Dixon

Published online: 08 Jan 2018.

Submit your article to this journal ⍈

View related articles ⍈

View Crossmark data ⍈

Routledge
Taylor & Francis Group

A Review of Manual and Computational Approaches for the Study of World Music Corpora

Maria Panteli, Emmanouil Benetos and Simon Dixon

Centre for Digital Music, Queen Mary University of London, UK

ABSTRACT

The comparison of world music cultures has been a recurring topic in the field of musicology since the end of the nineteenth century. Recent advances in technology in the field of Music Information Retrieval allow for a large-scale analysis of music corpora. We review manual and computational approaches in the literature that fall within the scope of music corpus research and world music analysis. With a large-scale computational music corpus analysis in mind, we compare the tools and research questions addressed by each study and discuss strengths and weaknesses. Taking into account critical remarks from experts in the field and challenges involved in a large-scale computational analysis, we discuss how this line of research can be improved in future work.

1. Introduction

The comparison of world music cultures has been of interest to musicologists and anthropologists since the end of the nineteenth century. Comparative musicology is termed as a subdiscipline of musicology (Adler, 1885) and experts in the field have made great progress in music data collection and analysis (Lomax, 1976; Savage, Merritt, Rzeszutek, & Brown, 2012; von Hornbostel & Sachs, 1961). Though traditional forms of musicological analysis provide a great deal of expert knowledge, the manual annotation involved in the process is time-consuming and limits the potential for large-scale insights.

The use of computers for the comparison and classification of music cultures was already conceived in the middle of the twentieth century (Bronson, 1949; Rhodes, 1965). Today, the advances of technology in the field of Music Information Retrieval (MIR) (Downie, 2003) allow for a thorough computational analysis of large music collections. The application of MIR techniques for the study of world music falls under the subdiscipline of Computational Ethnomusicology (Tzanetakis, Kapur, Schloss, & Wright, 2007). Several research projects have focused on the development of MIR tools for the study of specific world music corpora (Abdallah et al., 2017; Fillon et al., 2014; Kroher, Díaz-Báñez, Mora, & Gómez, 2016; Marolt, Vratanar, & Strle, 2009; Moelants, Cornelis, & Leman, 2009; Serra, 2011). Applications of MIR tools to the study and comparison of large world music corpora, however, are yet to be explored.

The fields of ethnomusicology and MIR have set the grounds for a large-scale comparison of world music. These fields bring different expertise to the challenging study of world music and the collaboration between the two has been considered a great advantage (Neubarth, Bergeron, & Conklin, 2011; Inskip & Wiering, 2015; van Kranenburg et al., 2010). We are interested in a large-scale computational analysis of world music integrating knowledge from both ethnomusicology and MIR.

With this paper, we review music corpus studies from both fields. We compare the size and scope of music corpora used in manual and computational approaches and contrast the research questions and findings of each approach. We also compare the music descriptors and data mining tools used in each study. Major comparative studies have also received criticism (Clarke, 2014; Fink, 2013; Nettl, 1970; Underwood, 2015). We highlight the strengths and weaknesses of state of the art research and point towards remaining challenges and lessons learnt for an improved computational study in the future.

This paper is organised as follows. Section 2 starts with a brief explanation of the terminology (Section 2.1) before moving to the literature review of music corpus studies with manual (Section 2.2) and computational (Section 2.3) approaches. Section 3 summarises criticism of major music corpus studies and Section 4 provides an overview of the challenges involved in large-scale computational analysis of world music. Further discussion of strengths and weaknesses of the state of the art research as well as concluding remarks are presented in Section 5.

CONTACT Maria Panteli ✉ m.panteli@qmul.ac.uk 💬 Maria Panteli, Centre for Digital Music, Queen Mary University of London, UK

2. Music corpus studies

There are numerous manual and computational approaches for the comparison of world music cultures via data mining of large music corpora. Studies reviewed here are selected based on four primary criteria. The first two criteria are: (a) the size of the corpus they analyse and (b) the research question they address. In particular, we exclude computational studies whose research question is not targeted at understanding the corpus itself but rather at assessing the methods or pursuing a specific algorithmic challenge. Likewise, we exclude manual studies that explore a relatively small (less than 100 recordings) or very specific corpus as it is unlikely that the methods are scalable and generalisable to non-specific corpora. The other two criteria require that (c) the studies under review are primarily concerned with the comparison of music cultures and (d) they provide a rather systematic approach in their methodology. Our primary interest is the comparison of world music cultures but since not many studies have considered a world music corpus we expand our review to include comparative music studies on popular, classical and folk and traditional music repertoires. Our review is primarily focused on studies that process music information from the sound recording or the music notation. World music studies based on historical, cultural or other metadata information (Baily & Collyer, 2006; Barrett, 1996) are beyond the scope of the present review.

While we attempt to list the most important corpus-based music studies in the literature so far we acknowledge that our list is not exhaustive and that we might have overlooked studies with similar methodologies but slightly different scope. A summary of the music studies under review along with their musical material and findings is presented in Tables 1 and 2.

2.1. Terminology

Terms and concepts frequently used in this literature review are explained in the paragraphs below. One of the most ambiguous terms is that of *world music*. The term can have various interpretations, and throughout the literature it has been used to denote popular and classical musics from around the world and from different eras (Bohlman, 2002). We use *world music* to define folk and traditional music from around the world. *Folk music* can be defined as

> the product of a musical tradition that has been evolved through the process of oral transmission. The factors that shape the tradition are: (i) continuity which links the present with the past; (ii) variation which springs from the creative impulse of the individual or the group; and (iii) selection by the community, which determines the

form or forms in which the music survives. (International Folk Music Council, 1955)

The folk and traditional music corpora we are interested in include Western folk music but exclude Western art music. *Music corpus* in this case defines a collection of music pieces in recorded form or musical notation. A *corpus-based study* addresses primarily research questions regarding the characteristics of the music corpus.

We often refer to two major research fields, ethnomusicology and MIR. *Ethnomusicology* traditionally focused on the study of non-Western music of oral traditions but today expands to the study of all music of the world in its social and cultural context (Dahlig-Turek, Klotz, Parncutt, & Wiering, 2012; Pegg, Myers, Bohlman, & Stokes, 2001). The term ethnomusicology was adopted to replace comparative musicology, but its concept is not only to study the world's musics from a comparative perspective but also to expand on the role of music within a culture and as a reflection of culture (Nettl, 2005). Another related field of musicology is *systematic musicology* (Adler, 1885), which includes the study of collections of music using analytical, statistical or computational approaches (Leman, 2008). Systematic musicology addresses 'how music practices can be understood, explained as a system (both from a psychoneuronal and social point of view), and possibly further explored and exploited (for example in connection with technology)' and it involves the study of music with methods from interdisciplinary fields including the humanities and sciences (Leman, 2008).

MIR is foremost concerned with the extraction and inference of musically relevant features (from the audio signal, symbolic representation or external sources such as web pages), indexing of music using these features, and the development of different search and retrieval schemes (for instance, content-based search, music recommendation systems or user interfaces for browsing large music collections) (Schedl, Gomez, & Urbano, 2014). *Digital musicology* is defined as interdisciplinary music research which encourages the use of technical infrastructure for musicology (Wiering & Benetos, 2013). A related term, *computational musicology*, has been used to denote the research area that combines questions, methods and insights from both musicology and computer science (Bel & Vecchione, 1993; Volk, Wiering, & van Kranenburg, 2011). The application of computational approaches to address musicological questions contributes to solving traditional problems, such as the time consuming task of manual music annotation, and opens new directions for musicological research, for example, big music data analyses.

We also make the following distinctions. The medium of music representation studied in the various manual and computational approaches reviewed in this study

can be either the *sound recording* or *music notation*. The former captures an acoustic representation of music as an audio signal, whereas the latter defines a symbolic representation of music as a score or other music notation system.

The systematic description of music can be made with either *manual annotations* or *automatically extracted features*. The former denotes the process of human experts manually annotating music attributes for each music piece, for example the Cantometrics and Cantocore system for world music (Lomax, 1976; Savage et al., 2012) and the Music Genome project for Western popular music (Prockup et al., 2015). Automatic feature extraction denotes the computational approach to derive music attributes from the audio signal, for example using the Librosa software (McFee et al., 2015), MARSYAS (Tzanetakis & Cook, 2000), MIR Toolbox (Lartillot & Toiviainen, 2007), or from the music notation, for example via the jSymbolic toolbox (McKay, 2010), or music21 (Cuthbert & Ariza, 2010). We refer to studies based on human annotations to music description as *manual approaches* and studies based on automatically extracted features as *computational approaches*. Manual approaches could still employ computational methods at a later stage of the analysis. However, the initial music annotation (human or automatic) that the analysis is based on, is what defines the approach as manual or computational throughout this review.

2.2. Manual approaches

Many studies in the field of ethnomusicology have considered and discussed the comparison of music cultures (Feld, 1984; Nettl, 2015; Nettl & Bohlman, 1991; Tenzer, 2006). Feld (1984) reflects on the approaches of comparative music studies and discusses the need for a qualitative comparison as well as the research questions that could contribute to the understanding of socio-musical practices. Tenzer (2006) explores music repertoires from around the world and reviews the contexts of their performance and creation and the ways to hear and conceive the different musical attributes. Nettl and Bohlman (1991) discuss the methodological and theoretical foundations as well as significant issues in the history of ethnomusicology. Nettl (2015) provides an overview of ethnomusicological research and focuses on concepts and issues that have caused a long ethnomusicological discourse.

A review of comparative music studies is also presented by Savage and Brown (2013). The authors redefine the field of comparative musicology, revisiting the research goals and discussing potential contributions of the field to the study of music classification, cultural evolution, human history, music universals and biological evolution. In this paper, we review a subset of these studies matching the criteria defined in the first paragraph of Section 2 and expand on music studies with computational approaches.

2.2.1. Audio recordings

One of the major comparative musicologists in the 1960s was Alan Lomax who collected more than 4000 recordings from many geographical areas and developed an annotation system, 'Cantometrics' (Lomax, 1976), to categorise the music cultures of the world (Lomax, 1968). Using a phylogenetic analysis, Lomax (1980, p. 39) observed two evolutionary roots, the Siberian and African Gatherers music styles. More recently, Savage and Brown (2014) analysed 259 traditional songs from 12 indigenous populations from Taiwan using 26 features from the 'Cantocore' system (Savage et al., 2012) focusing on rhythm, pitch, texture, and form. Using clustering analysis Savage and Brown (2014) showed that songs can be grouped in five clusters correlated with geographical factors and repertoire diversity. With a smaller corpus of 72 songs, Mora, Gómez, Gómez, Escobar-Borrego, and Díaz-Báñez (2010) developed a set of manual annotations for two flamenco styles, deblas and martinetes, and measured inter- and intra-style similarity with Euclidean distances and phylogenetic trees. A related study (Kroher, Gómez, Guastavino, Gómez, & Bonada, 2014) investigated similarity measures based on manually annotated and computationally extracted flamenco features and compared these measures to human ratings of melodic similarity.

Another application of comparative musicology is in the search for musical universals, i.e. the systematic comparison of the world's musics in order to understand how music features evolve over time and space (Brown & Jordania, 2011). The study of music universals received considerable attention in the 1970s with two journals, Ethnomusicology (1971) and The World of Music (1977), devoted to this topic. Savage et al. (2015) analysed 304 recordings contained in the 'Garland Encyclopedia of World Music' (Nettl, Stone, Porter, & Rice, 1998) using 32 features from the Cantocore and Cantometrics systems and instrument classification attributes as defined by von Hornbostel and Sachs (1961). Using phylogenies to control for historical relationships, continuous Markov processes to model rate of change and correlations of features across cultures, they were able to show that there are no *absolute* music universals but rather *statistical* universals. For example, there are 18 music features shared amongst many music cultures of the world and a network of 10 features that often occur together.

Other music comparative studies have focused on contrasting music to genetic and language evolution. Rzeszutek, Savage, and Brown (2012) annotated 421

traditional songs from 16 Austronesian-speaking populations from Taiwan and the northern Philippines using the Cantocore system. Correlations between music and genes showed that the majority of musical variability is due to differences within populations rather than differences between populations. In a similar study with 220 traditional songs from 9 indigenous populations from Taiwan, and a set of 41 descriptors (26 from Cantocore and 15 from Cantometrics systems), Brown et al. (2014) showed that population structures for genetics indicate stronger parallels to music than to language. Savage et al. (2015) compared genetic and musical diversity by analysing 680 traditional songs from two Ainu and 33 East Asian and circumpolar populations. The distribution of stylistic song types in music was similar to the distribution of DNA types and consistent with a 'triple structure' model of Japanese archipelago history. Le Bomin, Lecointre, and Heyer (2016) analysed 700 recordings from 58 patrimonies of rural areas in Gabon using 322 features on repertoire, form, instrument, metre, rhythm and melody. A phylogenetic analysis of repertoires showed that there is a predominant vertical transmission of musical characteristics such as metre, rhythm, and melody, where vertical transmission refers to the inheritance from ancestors in contrast to horizontal exchange from neighbours.

2.2.2. Music notation

A few studies were found using manual approaches to explore relatively large corpora of music notation. Bronson (1950) analysed several melodic, rhythmic and structural attributes of 100 British folk tunes from the 16th to the twentieth century. His findings include comparative statistics of the use of tune length, modes, meters, cadences and phrase patterns over the time span of five centuries. A related study (Savage, 2017), analysed 4125 British–American narrative songs from the Child ballads collection (Bronson, 1972) notated between 1575 and 1972. Hypotheses related to music culture evolution were tested and analysis showed that, amongst others, 'functional notes are more resistant to change than ornamental notes and substitutions are more likely to occur between small melodic distances than large ones' (Savage, 2017, p. 68). Freeman and Merriam (1956) compared the use of pitch intervals in a small corpus of 40 songs from two music cultures, the Ketu cult of Bahia, Brazil, and the Rada cult of Trinidad. He found that the two cultures can be distinguished by characteristic uses of major second and minor third intervals.

Volk and van Kranenburg (2012) developed an annotation method for 360 Dutch folk melodies including features capturing aspects of contour, rhythm and motif similarity. They found that the recurrence of charac-

teristic motifs is the most important feature for establishing similarity in Dutch folk melodies. Musicological hypotheses were also tested in a study of harmonic usage in American popular music as it evolved from the 1950s to the 1990s (Burgoyne, Wild, & Fujinaga, 2013). The authors used 1379 songs from the Billboard dataset with chord transcriptions manually annotated by experts (Burgoyne, Wild, & Fujinaga, 2011), and performed compositional data analysis to illustrate changes in harmonic usage over time. They found that there is a greater use of minor tonalities over time and dominant chords become less frequent than tonic and subdominant chords in recent songs.

A number of studies that have explored statistical techniques for the analysis of specific music notation corpora can be reviewed in Nettheim (1997), Temperley and Van Handel (2013), Gustar (2014), Walshaw (2014) and references therein. The majority of these studies focus on either small corpora or corpora and methods of very specific music styles and are thus beyond the scope of this review. It is also worth noting here that many world music cultures are orally transmitted and the resources of music notation are often limited. What is more, the study of music notation corpora employed computational tools from an early stage (Bronson, 1949; Scherrer & Scherrer, 1971) and therefore these are summarised under computational approaches in Section 2.3.2.

A summary of the manual approaches reviewed above is shown in Table 1.

2.3. Computational approaches

The use of computers for the comparison or classification of music cultures has been considered as early as the middle of the twentieth century (Bronson, 1949; Rhodes, 1965). Music corpus studies using computational tools have been considered in the fields of MIR and digital musicology. In these studies, the corpus is usually larger due to the efficiency of computational analysis but questions are raised on how representative and meaningful the automatically extracted features are. Below we review computational approaches using sound recordings and music notation.

2.3.1. Audio recordings

A number of computational approaches have focused on studying stylistic characteristics as they evolve over time. A study of 1010 recordings from the top 40 of the Billboard Hot 100 charts between 1965 and 2009 revealed that popular recordings became longer in duration and more sad-sounding over time (Schellenberg & von Scheve, 2012). Serrà, Corral, Boguñá, Haro, and Arcos (2012) analysed pitch, timbre, and loudness in 464,411

Table 1. The size of corpus, type of features and findings of music corpus studies in the literature using manual approaches. The count of descriptors is denoted by '–' when the exact number of features was not explicitly stated in the corresponding published work.

Study		Corpus size and description		Descriptor count and type	Main findings
Manual approaches – Audio recordings					
Lomax (1968)	1800	Various world music cultures	36	Cantometrics	Two possible evolutionary roots, the Siberian and African Gatherers music styles
Rzeszutek et al. (2012)	421	Traditional music from Taiwan and Philippines	26	Cantocore	The majority of musical variability is due to differences within populations rather than between populations
Savage and Brown (2014)	259	Traditional music from Taiwan	26	Cantocore	Songs grouped in 5 clusters correlated with geographical factors and repertoire diversity
Brown et al. (2014)	220	Traditional music from Taiwan	41	Cantocore and Cantometrics	Population structures for genetics indicate stronger parallels to music than to language
Savage et al. (2015)	680	Traditional songs from Japan and East Asia	41	Cantocore and Cantometrics	Stylistic song-types in music have similar distribution to DNA types and are consistent with a 'triple structure' model
Savage et al. (2015)	304	World music from the Garland encyclopedia	32	Cantocore and Cantometrics	No *absolute* music universals but *statistical* universals (18 common features and a network of 10 co-occurring features)
Le Bomin et al. (2016)	700	Traditional music from Gabon	322	Form, instrument, rhythm, melody	Predominant vertical transmission of musical characteristics such as metre, rhythm, and melody
Manual approaches – Music notation					
Bronson (1950)	100	British folk tunes	–	Mode, meter, form, cadence	Comparative statistics of the use of musical attributes over 5 centuries
Volk and van Kranenburg (2012)	360	Dutch folk songs	–	Melodic contour, rhythm, motif	The recurrence of characteristic motifs is the most important feature for establishing melodic similarity in Dutch folk melodies
Burgoyne et al. (2013)	1379	American popular music from the Billboard set	–	Beat-aligned chord transcriptions	There is a greater use of minor tonalities over time and dominant chords are less frequent than tonic and subdominant chords
Savage (2017)	4125	British–American narrative songs	–	Melodic contours	Functional notes are more resistant to change than ornamental notes and substitutions are more likely to occur between small melodic distances than large ones

recordings (between 1955 and 2010) of Western popular genres from the Million Song Dataset (MSD) (Bertin-Mahieux, Ellis, Whitman, & Lamere, 2011). Analysing music trends over the years revealed that more recent music shows less variety in pitch transitions, consistent homogenisation of the timbral palette and louder and potentially poorer volume dynamics. Shalit, Weinshall, and Chechik (2013) used 24,941 songs by 9222 artists (between 1922 and 2010) from the Million Song Dataset, audio features related to pitch, timbre and loudness, and topic models (Blei & Lafferty, 2006) and showed that the most influential songs were more innovative during the early 1970s and the mid 1990s than at other times. Mauch, MacCallum, Levy, and Leroi (2015) studied harmonic and timbral content in 17,094 songs covering 86% of the US Billboard Hot 100 between 1960 and 2010. Using topic modelling and clustering analysis they concluded that USA pop music evolved with particular rapidity during three stylistic 'revolutions' around 1964, 1983 and 1991.

With respect to non-Western music repertoires, Moelants et al. (2009) studied pitch distributions in 901 recordings from Central Africa.[1] They observed that

music from Central Africa does not conform to the 12-tone equal temperament, however in recent recordings there seems to be a tendency to the use of more equally tempered scales. Gómez, Haro, and Herrera (2009) studied music style classification in a dataset of 5905 recordings of Western and non-Western traditions using tonal, timbral, and rhythmic features. Their analysis verifies that Western music is more equal-tempered than non-Western and an investigation of which features correlate most with geographical regions indicated that latitude is mostly associated with tonal features and longitude with rhythmic ones. Other approaches to non-Western music analysis include the automatic classification of audio recordings into global cultural areas (Kruspe, Lukashevich, Abeßer, Großmann, & Dittmar, 2011; Zhou, Claire, & King, 2014), classification of ethnomusicological recordings by timbre features (Fourer, Rouas, Hanna, & Robine, 2014), the study of pitch distributions in Turkish (Bozkurt, 2008), Byzantine (Panteli & Purwins, 2013), and Indian classical (Ganguli, Gulati, Serra, & Rao, 2016) music, rhythmic patterns in Turkish (Holzapfel & Stylianou, 2009) and Indian art (Srinivasamurthy, Holzapfel, & Serra, 2014) music, and

---

[1] The Royal Museum for Central Africa http://music.africamuseum.be.

2.3.2. Music notation

Computational approaches have also been applied to analyse music in symbolic representation. A study of melodic contours from 6251 European folk songs from the Essen Folksong Collection (Schaffrath, 1995) revealed that melodies tend to exhibit an arc-shaped pitch contour (Huron, 1996). Zivic, Shifres, and Cecchi (2013) analysed classical music scores between 1700 and 1930 from the Peachnote corpus (Viro, 2011) which consists of more than 900,000 scores. By studying bigrams of melodic intervals they were able to show that classical music styles are distinguished by characteristic differences in their distribution of melodic intervals over time. Pamjav, Juhász, Zalán, Németh, and Damdin (2012) analysed pitch sequences of 31 Eurasian and North American folksong collections, each of them consisting of 1000–2500 melodies. Using Self Organising Maps (SOMs) and Multi-Dimensional Scaling approaches they showed that there is a significant correlation between population genetics and folk music, and that maternal lineages in folk music are more prominent than paternal lineages. Volk and de Haas (2013) studied syncopation in ragtime music by analysing melodic patterns from 11,000 ragtime MIDI files. The authors confirmed the musicological hypothesis that the use of tied syncopations increased in the ragtime era after 1902 in comparison to the use of untied syncopations.

Aarden and Huron (2001) analysed the phrase endings from European folk melodies of the Essen Folksong Collection. From a total of approximately 950 melodies they observed that Western European melodies are more likely to have their melodies ending on the tonic than Eastern European melodies. Juhász (2006) studied melodic contours of approximately 9000 folksongs from Slovak, French, Sicilian, Bulgarian, English and Hungarian music cultures. Using SOMs it was shown that a common set of contour types was shared amongst the six cultures and that these contour types are represented especially in the Hungarian and Slovak traditions. In a subsequent study including music from additional cultures of Eurasia, SOM analysis revealed that the use of melodic contours in different geographical areas can be grouped into two main clusters (Juhász, 2009). Shanahan, Neubarth, and Conklin (2016) analysed 2083 folksongs from the Frances Densmore's collection of Native American music using attributes from the jSymbolic set (McKay, 2010) and information-theoretic measures. Contrast mining methods (Dong & Li, 1999) were employed to compare music in different social contexts.

Their analysis showed, amongst others, that nature songs have low variability of events, love songs have larger melodic intervals and higher pitch registers, and war and dance songs are 'high arousal' songs but on opposite ends of the valence spectrum on Russell's Circumplex model (Russell, 2003). Other approaches to studying music corpora include the classification of folk Dutch melodies with local and global features (van Kranenburg, Volk, & Wiering, 2013), and the analysis of melodic patterns in Cretan folk songs (Conklin & Anagnostopoulou, 2011).

A summary of the computational approaches reviewed above is shown in Table 2.

3. Critical remarks

Music corpus studies seen in the literature so far have received considerable criticism. In this section, we review issues raised about the most popular comparative studies.

The work by Lomax (1976) has concerned ethnomusicologists and anthropologists (Dubinskas, 1983; Feld, 1984; Nettl, 1970). Some of the critical remarks as Nettl (1970) suggests are that the dataset samples too few songs from each culture and that the annotation system (Cantometrics) may not be representative because the annotators may lack a complete understanding of the music: 'Can someone understand a music without immersing oneself in it for years?' (Nettl, 1970, p. 439). Furthermore, annotations may not be very reliable due to the difficulty of the task for human listeners,

> evaluating by ear such elusive qualities as vocal rasp, nasality, and vocal width (which are not standard or widely used concepts in musicology) and assigning their relative degree in a recording according to a scale of up to ten points would appear to be a questionable procedure. (Nettl, 1970, p. 440)

Feld (1984) discusses the need for a qualitative and intensive comparative musicology and comments that

> the best way to answer Lomax's questions about the systematic nature of musical representation in social organisation is to study them on the ground, in the field, up close, over long periods of time, where sound structures are observably and undeniably socially structured.

He also defines research questions under six domains (competence, form, performance, environment, value and equality, theory) that could contribute to the comparison of socio-musical realities and practices.

Savage and Brown (2014) described key themes in comparative musicology and included, amongst others, the generation of a musical map of the world reflecting aspects of cultural diversity and evolution. Clarke (2014) criticises the properties of the music to be considered in the creation of such a map: 'Should it be based on musical production (composition, performance), or consumption (concert going, private listening)? Should we

Table 2. The size of corpus, the type of features, and the findings of music corpus studies in the literature using computational approaches. The count of descriptors is denoted by '−' when the exact number of features was not explicitly stated in the corresponding published work.

Study	Corpus size and description		Descriptor count and type		Main findings
Computational approaches – Audio recordings					
Moelants et al. (2009)	901	Traditional music from Central Africa	1200	pitch histogram	African music does not conform to the fixed chromatic scale but recent recordings use more equally-tempered scales
Gómez et al. (2009)	5905	Traditional music from Western and non-Western countries	23	timbre, rhythm, tonality	Western music is more equal-tempered than non-Western, latitude mostly associated with tonal features and longitude with rhythmic ones
Schellenberg and von Scheve (2012)	1010	American popular music from the Billboard set	4	tempo, mode, duration, gender of vocalist	Popular recordings became longer in duration and more sad over time
Serrà et al. (2012)	464,411	Western popular music from the Million Song Dataset	–	pitch, timbre, loudness	Recent music shows less variety in pitch transitions, consistent homogenisation of the timbral palette, and louder and potentially poorer volume dynamics
Shalit et al. (2013)	24,941	Western popular music from the Million Song Dataset	–	pitch, timbre, loudness	The most influential songs were more innovative during the early 1970's and the mid 1990's
Mauch et al. (2015)	17,094	American popular music from the Billboard set	–	tonal and timbral topics	Pop music evolved with particular rapidity during three stylistic revolutions around 1964, 1983 and 1991
Computational approaches – Music notation					
Huron (1996)	6251	European folk music from the Essen Folksong Collection	–	melodic contours	European folk melodies tend to exhibit an arch-shaped pitch contour
Aarden and Huron (2001)	950	European folk music from the Essen Folksong Collection	–	melodic phrase ending	Western European melodies are more likely to have their melodies ending on the tonic than Eastern European melodies
Juhász (2006)	9000	folk music from six European cultures	–	melodic contours	There is a common set of contour types shared amongst the six cultures and these contour types are represented especially in the Hungarian and Slovak traditions
Juhász (2009)	19,733	folk music from 11 Eurasian cultures	–	melodic contours	The use of melodic contours in different geographical areas can be grouped into two main clusters
Pamjav et al. (2012)	≈ 31000	folk music from Eurasia and North-America	–	melodic contours	There is a significant correlation between population genetics and folk music and maternal lineages in folk music are more prominent than paternal lineages
Zivic et al. (2013)	≈ 900, 000	Western classical music	–	melodic interval bigrams	Classical music styles show characteristic differences in their distribution of melodic intervals over time
Volk and de Haas (2013)	11,000	Ragtime music	–	syncopation in melodic patterns	The use of tied syncopations in comparison to untied syncopations is increased in the ragtime era after 1902
Shanahan et al. (2016)	2083	Native American folk music	18	pitch, duration, intervals, onsets	Contrasts between musics in different social contexts, e.g. nature songs have low variability of events, love songs have larger melodic intervals

consider the public sphere (larger, widely advertised events) or the (semi-)private (domestic get-togethers and community gatherings)?' (Clarke, 2014, p. 9). He also raises a point about temporal evolution, 'traditions evolve, styles mutate, patterns of consumption change', that is not captured in a static collection of music and a projected local map 'would be just one snapshot on a much larger diachronic continuum' (Clarke,2014, p. 9).

Large-scale computational approaches to music corpus analysis have also received criticism. One of the major issues for the study of Serrà et al. (2012) is the suitability

of the corpus. Fink (2013) observes that the study investigates evolutionary trends in the Million Song Dataset, a data-set created primarily for the evaluation of MIR algorithms. As Fink (2013) mentions, 'any conclusions drawn from the MSD are already constrained by the assumptions and mindset of the industry-research teams that created the database'. Another major drawback is that the music coding system is not easily interpretable and numerical representations derived from the model can be questioned as to whether they contain meaningful musical information (Fink, 2013; Wallmark, 2013). What

is more, Western bias may influence the interpretation of results (Fink, 2013) and the social context in which the music is actually heard is disregarded in such computational analysis (Wallmark, 2013).

Similar critical remarks apply to the study by Mauch et al. (2015). Underwood (2015) discusses whether measures of stylistic 'distance' between songs can indicate cultural change and how robust these measures can be. In another post, Underwood, Long, So, and Zhu (2016) suggest that statistical significance is calculated in a misleading way, 'only two of the three "revolutions" it (Mauch et al., 2015) reported are really significant at $p < 0.05$, and it misses some odd periods of stasis that are just as significant as the periods of acceleration'. Thompson (2015) points to some alternative factors, namely the change in the Billboard measurement system in 1991, that might have contributed to observing a music revolution in 1991 as concluded by Mauch et al. (2015).

The critical remarks presented above for different studies in the literature are often overlapping. For example, the suitability of the corpus has been questioned in both manual (Lomax, 1976) and computational (Serrà et al., 2012) approaches. The reliability of music annotations can be an issue in both approaches (see e.g. remarks by Nettl (1970) and Fink (2013) above). On the one hand human experts may not be able to reliably annotate fine-grained musical characteristics judging solely by ear. On the other hand, computational systems may fail to capture high-level attributes for example aspects of music perception and cognition. The above criticism gives valuable feedback on challenges that need to be addressed for improved music corpus-based studies.

4. Challenges for a large-scale computational analysis of world music

In the literature reviewed above, we have seen comparative approaches of world music with relatively small corpora of audio recordings (Lomax, 1980; Savage et al., 2015), and large-scale approaches focusing mainly on Western music corpora (Mauch et al., 2015; Serrà et al., 2012). A large-scale comparative study of world music cultures has not been addressed yet. Nettl (2005, p. 11) reflects on the interculturally comparative perspective of ethnomusicologists who

> look at each musical culture from a viewpoint that relates it to the world of music, a world comprised of a multitude of musical cultures that are alike in some ways and different in others, and they believe that insight can be gained from comparison.

We too value the insights gained by such a comparison and we are particularly interested in studying the ways

musical cultures of the world are different or alike and the reasons for such differences and similarities.

Other research questions that could be addressed with a music comparison include identifying which aspects of music are universal to all cultures and investigating whether language influences the musical tradition. Computational tools could aid such comparisons and large-scale analysis could increase the impact of any findings. Large-scale approaches can also be useful towards analysing the music of a specific culture or style, by simply making the conclusions more reliable or by enabling the study of trends over time or smaller geographical regions. However, a large-scale comparison with computational tools includes several challenges with respect to processing information from the metadata and the audio recordings as well as generalising findings from big data collections. Below we list the major challenges associated with this line of research.

Restricted access to audio recordings. While several research projects and institutions make great efforts to increase the accessibility to audio music collections (Abdallah et al., 2017; Franzen, 2016; Porter, Sordo, & Serra, 2013), a lot of recorded world music is still not available for research due to copyright and other ethical issues.[2] To create a world music corpus, we need to combine sound recordings from distributed sources and collections. This brings up further challenges in setting up legal agreements with the owners of each collection and processing the information from each source in a unified manner. The Digital Music Lab project (Abdallah et al., 2017) proposed to circumvent this problem by performing the analysis locally on each collection and aggregating the results centrally.

Unbalanced collections. Access to fieldwork in ethnomusicology as well as in other ethnographic research is affected by spatial and temporal parameters (Barz & Cooley, 2008; Hammersley, 2006). In large collections of world music recordings, it is often the case that Western-influenced music traditions are more represented than non-Western. A comparative study on world music however requires a balanced corpus with a good representation of the geographical and cultural diversity of world music as well as a good temporal spread of the music eras.

Corpus creation. Creating a corpus suitable for the computational study of world music imposes further challenges in terms of qualitative and quantitative criteria. As seen in past criticism (Section 3), the corpus needs to include the most representative samples from each music culture (Nettl, 1970), and the assumptions

---

[2]In some cases, copyright exceptions encourage research with audio recordings as long as the research is non-commercial, the resources are properly acknowledged, and the research results cannot recreate the original works (see e.g. regulations for research in the UK at https://www.gov.uk/government/organisations/intellectual-property-office).

made to create the corpus must be in line with the research questions under study (Fink, 2013). This requires addressing what defines a good sample, how to balance the diversity, and how to maximise the size of this corpus to obtain large-scale results. Serra (2014) defines five criteria, namely the purpose, coverage, completeness, quality, and reusability to be taken into account when creating corpora for the computational study of music. Similar criteria are also followed by Kroher et al. (2016) for the creation of a corpus for the computational study of flamenco music.

Interpretation of metadata. In order to study the relationships between musical content and metadata of world music, spatio-temporal information of the origins of the music is required. In world music recordings, the temporal information associated with the metadata represents the time the music was recorded but does not necessarily represent the time at which it was composed. For example, for most folk music the time and location of a song's composition remains unknown. What is more, unlike Western popular music where there is often a common agreement concerning the taxonomy of music styles, in world music the classification of music styles is still in great discourse (Clayton, Herbert, & Middleton, 2003; Lomax & Berkowitz, 1972). The assumptions made when creating the metadata need also to be considered, for example, the purpose of the metadata creation and the background and interest of the curators. There are, therefore, greater challenges involved in processing the metadata for world music.

Incorrect metadata. Depending on the collector and the era in which a recording session took place, the information registered for each recording varies vastly or is absent altogether. A great challenge is, therefore, to combine all the available information and create a consistent database of metadata. In several cases, information on the culture or language of a recording is misspelled or the registered location is inconsistent with the latest geopolitical maps (e.g. 'USSR' or 'Yugoslavia' whose borders and political status have changed). Automatic correction of this type of metadata requires techniques from natural language processing and geopolitical database matching.

Lack of ground truth. The comparison of world music cultures comprises an exploratory type of research. There is scattered information concerning the ways in which music cultures might be similar, but there is no single source defining all possible relations between them. For computational approaches, it is often necessary to have a ground truth which is used to train and also assess the performance of the algorithms. The notion of music similarity is subjective and considering especially the diversity in world music, creating a ground truth of music similarity judgements is very difficult. Not only

is the music diverse and the corpus large, but also music perception varies between listeners with different cultural backgrounds (Stevens, 2012).

Non-robust computational music processing. The automatic extraction of musical attributes is necessary for the large-scale computational analysis of world music. Several computational tools for the analysis of music signals have been designed for the primary aim of Western music analysis (Futrelle & Downie, 2002). This means that the tools may sometimes not be reliable for automatic processing of world music recordings and further developments should be considered. What is more, the extraction of music information from the audio signal can be largely affected by the audio recording quality (Urbano, Bogdanov, Herrera, Gómez, & Serra, 2014). This is especially a challenge in world music recordings where recording conditions vary vastly and material is preserved with different degrees of fidelity. The majority of world music recordings originate from fieldwork, where continuous audio streams need to be further segmented and curated (either manually or automatically). The evaluation of audio descriptors becomes an essential task in large-scale computational analysis (Panteli & Dixon, 2016).

Limitations of computational music content description. Music descriptors extracted automatically from the audio signal are unable to model properties the same as music descriptors extracted manually by human experts. Computational approaches can more accurately capture low-level characteristics of the audio signal whereas manual approaches can more reliably describe high-level features such as aspects of music perception and cognition. For example, an instrument classification system built for manual annotation referred to instrument properties like 'directly struck' and 'indirectly struck' idiophone (von Hornbostel & Sachs, 1961). In automatic instrument classification, algorithms are trained on features capturing low-level characteristics of the signal for example the 'zero-crossing rate' and 'Mel frequency cepstrum coefficients' (Aucouturier, Pachet, & Sandler, 2005) and higher level classification, such as by instrument type, is performed by learning mappings from the low-level to high-level features. The limitations of computational music description in capturing high-level music properties should be taken into account.

Missing context. The analysis of audio recordings from large music archives has great potential via the application of MIR and data mining technologies. However, information extracted solely from the audio signal is incapable of capturing all the aspects of the practice of a music tradition. Music context often lies beyond the audio signal and understanding this context requires

processing other forms of music representation not captured by the algorithms and tools reviewed in this study. The computational study of world music can benefit from the incorporation of additional musical context, for example, music notation, social context and experts' knowledge and analyses. For example, introducing a music ontology framework (Raimond, Abdallah, Sandler, & Giasson, 2007) covering aspects of world music could contribute significantly to the missing context of audio recordings.

Cultural bias. A cultural bias could affect many aspects of a particular study, from the point of acquiring and selecting data, which features to extract or annotate, which (mathematical, behavioural, computational, cognitive) model to use, and how to interpret the results. The risk of cultural bias is particularly high considering the study of world music requires knowledge of many different music traditions.

5. Discussion

As discussed in Section 4, we are interested in a large-scale comparative study of world music with computational tools. We reviewed comparative studies with manual and computational approaches and discussed the challenges involved in a large-scale study. Below we summarise our conclusions and directions for future work.

A large-scale comparison of world music cultures using computational tools has not been addressed yet. Computational approaches to music corpus analysis have mainly focused on Western popular music (Mauch et al., 2015; Serrà et al., 2012; Shalit et al., 2013). Computational approaches that have considered world music have either used a relatively small and geographically restricted corpus (e.g. less than 1000 recordings to compare African scales (Moelants et al., 2009) or aimed to answer different research questions (e.g., which audio features are most suitable for world music classification (Gómez et al., 2009; Kruspe et al., 2011). Manual approaches that focus on world music are usually restricted to relatively small data-sets (with the exception of Lomax (1976) and Savage (2017) analysing more than 4000 recordings the remaining approaches have studied corpora of less than 1000 recordings (Brown et al., 2014; Le Bomin et al., 2016; Rzeszutek et al., 2012; Savage & Brown, 2014).

The largest corpora in comparative music research have been considered in studies analysing music notation (e.g., almost a million scores were analysed to study the distribution of pitch intervals in classical music (Zivic et al., 2013). The advantage of music notation is that it is usually easier to access in contrast to copyrighted sound recordings which are often not available for research. However, while music structure is well represented in music notation, acoustic and performance-style characteristics are not captured. What is more, music notation does not exist in all world music cultures and different notation languages and formats across different styles make the comparison difficult. Therefore a world music comparison based on audio recordings is more plausible in this case.

Given the corpora and methods used in both manual and computational approaches to music corpus analysis as shown in Tables 1 and 2, and the corresponding criticism as explained in Section 3, we emphasise the following issues that need to be addressed for future computational studies.

The majority of the criticism for both manual and computational approaches has focused on the sample not being representative for the research question under investigation (see Fink (2013) for the review of Serrà et al. (2012)), the sample size not being large enough for statistical significance of the findings (see Nettl (1970) for the review of Lomax (1976)), and the sample not being inclusive of all music cultures of the world (see Clarke (2014) for a review of Savage and Brown (2014)). As discussed in Section 4, the sample size can be maximised by combining recordings from distributed sources and collections and sampling methods can be employed to balance the corpus. The selection criteria to ensure the collection is representative with respect to music style can be fulfilled if additional metadata are available, for example, the geographical origins, the language and culture of the performers, the year it was recorded or the era of the music it represents, as well as the primary purpose of the fieldwork study or recording collection.

Criticism of computational approaches raised the issue of the automatically extracted features not being suitable to capture meaningful music attributes (see Fink (2013) for a review of Serrà et al. (2012) and Underwood (2015) for a review of Mauch et al. (2015)). What is more, for both manual and computational approaches the set of music descriptors has been criticised for not being complete, i.e. not capturing all essential information about the music in comparison (see Underwood (2015) for a review of Mauch et al. (2015), Nettl (1970) for a review of Lomax (1976)). The audio features need to be perceptually evaluated or otherwise demonstrated to be meaningful and a thorough list of necessary music descriptors should be developed. An alternative solution could be to not rely solely on a set of features, e.g. derived from the music notation (where performance-specific characteristics are missing), or audio signal (where high-level or perceptual features are difficult to capture), but to combine both notation, audio, and metadata information for a more balanced study of world music. For example, semi-automatic approaches where manual annotations

complement automatically extracted features (Cabrera et al., 2008; van Kranenburg et al., 2010) could provide a better representation of the music that could also partly scale to larger corpora. In addition, approaches that learn from weakly labelled data (e.g. using metadata as weak labels) could also be used to extract more reliable high-level MIR features.

Large-scale music comparisons and evolutionary analyses require advanced computational methods. Extra care needs to be taken to not violate assumptions of the underlying statistical tests (see Underwood et al. (2016) for a review of Mauch et al. (2015)). What is more, a good understanding of the musical characteristics of the corpus is required by the person conducting the research to avoid biasing the methodology or the interpretation of any findings (see e.g. the Western bias remark by Fink (2013) in Section 3). Conclusions are more likely to be reliable if validated by experts in other disciplines including musicology, biology, statistics, history and anthropology.

The fields of musicology and MIR have set the grounds for large-scale music corpus studies. By reviewing manual and computational approaches we highlighted the advantages and strengths of state of the art studies. Manual approaches benefit from direct expert knowledge but are limited by the time-consuming task of manual annotation. Computational approaches benefit from the efficient automatic music processing but can be limited by the knowledge represented in the derived attributes. Criticism of popular music corpus studies focuses on the suitability and size of the corpus as well as how meaningful and robust the extracted music attributes are. Taking into account the challenges involved in a large-scale computational analysis of world music and the aforementioned critical remarks we discussed how music corpus studies can be improved in the future. We strongly believe that a large-scale computational comparison is now plausible and, if done properly, could provide valuable insights into world music.

Funding

ORCID

Emmanouil Benetos http://orcid.org/0000-0002-6820-6764

References

Aarden, B., & Huron, D. (2001). Mapping European Folksong: Geographical localization of musical features. *Computing in Musicology, 12*, 169–183.

Abdallah, S., Benetos, E., Gold, N., Hargreaves, S., Weyde, T., & Wolff, D. (2017). The Digital Music Lab: A big data infrastructure for digital musicology. *ACM Journal on Computing and Cultural Heritage, 10*(1). doi:10.1145/2983918

Adler, G. (1885). Umfang, Methode und Ziel der Musikwissenschaft. *Vierteljahresschrift für Musikwissenschaft, 1*(1), 5–20.

Aucouturier, J. J., Pachet, F., & Sandler, M. (2005). "The way it sounds": Timbre models for analysis and retrieval of music signals. *IEEE Transactions on Multimedia, 7*(6), 1028–1035. doi:10.1109/TMM.2005.858380

Baily, J., & Collyer, D. M. (2006). Introduction: Music and migration. *Journal of Ethnic and Migration Studies, 32*(2), 167–182.

Barrett, J. (1996). World music, nation and postcolonialism. *Cultural Studies, 10*(2), 237–247.

Barz, G. F., & Cooley, T. J. (Eds.). (2008). *Shadows in the field: New perspectives for fieldwork in ethnomusicology*. New York, NY: Oxford University Press.

Bel, B., & Vecchione, B. (1993). Computational musicology. *Computers and the Humanities, 27*(1), 1–5.

Bertin-Mahieux, T., Ellis, D. P. W., Whitman, B., & Lamere, P. (2011). The million song dataset. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 591–596). Miami.

Blei, D. M., & Lafferty, J. D. (2006). Dynamic topic models. In *Proceedings of the 23rd International Conference on Machine Learning (ICML 2006)* (pp. 113–120). Pittsburgh: ACM Press.

Bohlman, P. V. (2002). *World music: A very short introduction*. New York, NY: Oxford University Press.

Bozkurt, B. (2008). An automatic pitch analysis method for Turkish maqam music. *Journal of New Music Research, 37*(1), 1–13.

Bronson, B. H. (1949). Mechanical help in the study of Folk song. *Journal of American Folklore, 62*(244), 81–86.

Bronson, B. H. (1950). Some observations about melodic variation in British-American Folk tunes. *Journal of the American Musicological Society, 3*, 120–134.

Bronson, B. H. (1972). *The traditional tunes of the child ballads: With their texts, according to the extant records of Great Britain and America [4 Volumes]*. Princeton, NJ: Princeton University Press.

Brown, S., & Jordania, J. (2011). Universals in the world's musics. *Psychology of Music, 41*(2), 229–248.

Brown, S., Savage, P., Ko, M.-S., Stoneking, M., Ko, Y.-C., Loo, J.-H., & Trejaut, J. (2014). Correlations in the population structure of music, genes and language. *Proceedings of the Royal Society B-Biological Sciences, 281*(1774). doi:20132072

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2011). An expert ground-truth set for audio chord recognition and music analysis. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 633–638). Miami.

Burgoyne, J. A., Wild, J., & Fujinaga, I. (2013). Compositional data analysis of harmonic structures in popular music. In J. Yust, J. Wild, & J. A. Burgoyne (Eds.), *Mathematics and computation in music. MCM 2013*. Lecture Notes in Computer Science (Vol. 7937, pp. 52–63). Berlin, Heidelberg: Springer.

Cabrera, J. J., Díaz-báñez, J. M., Escobar-Borrego, F. J., Gómez, E., Gómez, F., & Mora, J. (2008). Comparative melodic analysis of a cappella flamenco cantes. In *Fourth*

Conference on Interdisciplinary Musicology (CIM08) (pp. 1–8). Thessaloniki.

Clarke, D. (2014). On not losing heart: A response to savage and Brown's "Toward a New Comparative Musicology". *Analytical Approaches to World Music, 3*(2), 1–14.

Clayton, M., Herbert, T., & Middleton, R. (Eds.). (2003). *The cultural study of music: A critical introduction*. New York, NY: Routledge.

Conklin, D., & Anagnostopoulou, C. (2011). Comparative pattern analysis of cretan folk songs. *Journal of New Music Research, 40*(2), 119–125.

Cuthbert, M. S., & Ariza, C. (2010). Music21: A toolkit for computer-aided musicology and symbolic music data. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 637–642). Utrecht.

Dahlig-Turek, E., Klotz, S., Parncutt, R., & Wiering, F. (2012). *Musicology (Re-) Mapped: Discussion Paper*. European Science Foundation. Retrieved from http://www.esf.org/fileadmin/Public_documents/Publications/musicology.pdf

Dong, G., & Li, J. (1999). Efficient mining of emerging patterns: Discovering trends and differences. In *Proceedings of the 5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 43–52). San Diego.

Downie, J. S. (2003). Music information retrieval. *Annual Review of Information Science and Technology, 37*(1), 295–340.

Dubinskas, F. A. (1983). A musical Joseph's coat: Patchwork patterns and social significance in world musics. *Reviews in Anthropology, 10*(3), 27–42.

Feld, S. (1984). Sound structure as social structure. *Ethnomusicology, 28*(3), 383–409.

Fillon, T., Simonnot, J., Mifune, M.-F., Khoury, S., Pellerin, G., Le Coz, M., ... Fourer, D. (2014). Telemeta: An open-source web framework for ethnomusicological audio archives management and automatic analysis. In *1st International Digital Libraries for Musicology Workshop (DLfM 2014)*. London. doi:10.1145/2660168.2660169

Fink, R. (2013). *Big (Bad) data*. Retrieved from http://musicologynow.ams-net.org/2013/08/big-bad-data.html

Fourer, D., Rouas, J.-L., Hanna, P., & Robine, M. (2014). Automatic timbre classification of ethnomusicological audio recordings. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 295–300).

Franzen, R. (2016). Europeana sounds: An interface into European sound archives. *Sound Studies, 2*(1), 103–106. doi:10.1080/20551940.2016.1154303

Freeman, L. C., & Merriam, A. P. (1956). Statistical classification in anthropology: An application to ethnomusicology. *American Anthropologist, 58*, 464–472.

Futrelle, J., & Downie, J. S. (2002). Interdisciplinary communities and research issues in Music Information Retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 215–221). Paris.

Ganguli, K. K., Gulati, S., Serra, X., & Rao, P. (2016). Data-Driven Exploration of Melodic Structures in Hindustani Music. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 605–611). New York, NY.

Gómez, E., Haro, M., & Herrera, P. (2009). Music and geography: Content description of musical audio from different parts of the world. In *Proceedings of the International*

Society for Music Information Retrieval Conference (pp. 753–758). Kobe.

Gustar, A. J. (2014). *Statistics in historical musicology* (PhD thesis). Open University, Milton Keynes.

Hammersley, M. (2006). Ethnography: Problems and prospects. *Ethnography and Education, 1*(1), 3–14. doi:10.1080/17457820500512697

Holzapfel, A. (2010). *Similarity methods for computational ethnomusicology* (Unpublished doctoral dissertation). University of Crete, Crete.

Holzapfel, A., & Stylianou, Y. (2009). Rhythmic similarity in traditional Turkish music. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 99–104). Kobe.

Huron, D. (1996). The melodic arch in Western folksongs. *Computing in Musicology, 10*, 3–23.

Inskip, C., & Wiering, F. (2015). In their own words: Using text analysis to identify musicologists' attitudes towards technology. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 455–461). Malaga.

International Folk Music Council. (1955). *Resolutions: Definition of folk music* (Vol. 7). Sao Paulo: International Council for Traditional Music.

Juhász, Z. (2006). A systematic comparison of different European folk music traditions using self-organizing maps. *Journal of New Music Research, 35*(2), 95–112.

Juhász, Z. (2009). Automatic segmentation and comparative study of motives in eleven folk song collections using self-organizing maps and multidimensional mapping. *Journal of New Music Research, 38*(1), 71–85.

Kroher, N., Díaz-Báñez, J.-M., Mora, J., & Gómez, E. (2016, May). Corpus COFLA: A research corpus for the computational study of flamenco music. *Journal on Computing and Cultural Heritage, 9*(2), 10:1–10:21. Retrieved from http://doi.acm.org/10.1145/2875428

Kroher, N., Gómez, E., Guastavino, C., Gómez, F., & Bonada, J. (2014). Computational models for perceived melodic similarity in a capella flamenco singing. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 65–70). Taipei.

Kruspe, A., Lukashevich, H., Abeßer, J., Großmann, H., & Dittmar, C. (2011). Automatic classification of musical pieces into global cultural areas. In *AES 42nd International Conference* (pp. 1–10). Ilmenau. doi:10.13140/2.1.4970.5606

Lartillot, O., & Toiviainen, P. (2007). A Matlab toolbox for musical feature extraction from audio. In *International Conference on Digital Audio Effects* (pp. 237–244). Bordeaux.

Le Bomin, S., Lecointre, G., & Heyer, E. (2016). The evolution of musical diversity: The key role of vertical transmission. *PLoS ONE, 11*(3), e0151570. doi:10.1371/journal.pone.0151570

Leman, M. (2008). Systematic musicology at the crossroads of modern music research. In *Systematic and comparative musicology: Concepts, methods, findings* (pp. 89–115). Frankfurt am Main: Peter Lang.

Lomax, A. (1968). *Folk song style and culture*. New Brunswick: American Association for the Advancement of Science.

Lomax, A. (1976). *Cantometrics: An approach to the anthropology of music*. Berkeley: University of California Extension Media Center.

Lomax, A. (1980). Factors of musical style. In S. Diamond (Ed.), *Theory & practice: Essays presented to gene weltfish* (pp. 29–58). The Hague: Mouton.

Lomax, A., & Berkowitz, N. (1972). The evolutionary taxonomy of culture. *Science, 177*(4045), 228–239. doi:10.1126/science.177.4045.228

Marolt, M., Vratanar, J. F., & Strle, G. (2009). Ethnomuse: Archiving folk music and dance culture. In *IEEE EUROCON 2009* (pp. 322–326). St.-Petersburg.

Mauch, M., MacCallum, R. M., Levy, M., & Leroi, A. M. (2015). The evolution of popular music: USA 1960–2010. *Royal Society Open Science, 2*(5), 150081.

McFee, B., McVicar, M., Raffel, C., Liang, D., Nieto, O., Battenberg, E., ... Holovaty, A. (2015). *Librosa: 0.4.1, Zenodo*. doi:10.5281/zenodo.32193. Retrieved from http://github.com/bmcfee/librosa/tree/0.4.1

McKay, C. (2010). *Automatic music classification with jMIR* (PhD thesis). McGill University, Canada.

Moelants, D., Cornelis, O., & Leman, M. (2009). Exploring African tone scales. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 489–494). Kobe.

Mora, J., Gómez, F., Gómez, E., Escobar-Borrego, F., & Díaz-Báñez, J. M. (2010). Characterization and melodic similarity of a cappella flamenco cantes. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 351–356). Utrecht.

Nettheim, N. (1997). A bibliography of statistical applications in musicology. *Musicology Australia, 20*(1), 94–106.

Nettl, B. (1970). Review of folk song style and culture by Alan Lomax source. *American Anthropologist, New Series, 72*(2), 438–441.

Nettl, B. (2005). The harmless drudge: Defining ethnomusicology. In B. Nettl (Ed.), *The study of ethnomusicology thirty-one issues and concepts* (2nd ed., pp. 3–15). Urbana and Chicago: University of Illinois Press.

Nettl, B. (2015). *The study of ethnomusicology: Thirty-three discussions* (3rd ed.). Champaign: University of Illinois Press.

Nettl, B., & Bohlman, P. V. (Eds.). (1991). *Comparative musicology and anthropology of music: Essays on the history of ethnomusicology*. Chicago: University of Chicago Press.

Nettl, B., Stone, R. M., Porter, J., & Rice, T. (Eds.). (1998). *The Garland encyclopedia of world music*. (1998–2002 ed.). New York, NY: Garland Pub.

Neubarth, K., Bergeron, M., & Conklin, D. (2011). Associations between musicology and music information retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 429–434). Miami.

Pamjav, H., Juhász, Z., Zalán, A., Németh, E., & Damdin, B. (2012). A comparative phylogenetic study of genetics and folk music. *Molecular Genetics and Genomics, 287*(4), 337–349.

Panteli, M., Benetos, E., & Dixon, S. (2016). Learning a feature space for similarity in world music. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 538–544). New York, NY.

Panteli, M., & Dixon, S. (2016). On the evaluation of rhythmic and melodic descriptors for music similarity. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 468–474). New York, NY.

Panteli, M., & Purwins, H. (2013). A quantitative comparison of chrysanthine theory and performance practice of scale tuning, steps, and prominence of the octoechos in byzantine chant. *Journal of New Music Research, 42*(3), 205–221. doi:10.1080/09298215.2013.827215

Pegg, C., Myers, H., Bohlman, P. V., & Stokes, M. (2001). Ethnomusicology. In Grove Music Online. Oxford Music Online. Retrieved from http://doi.org/10.1093/gmo/9781561592630.article.52178

Porter, A., Sordo, M., & Serra, X. (2013). Dunya: A system for browsing audio music collections exploiting cultural context. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 101–106). Curitiba.

Prockup, M., Ehmann, A. F., Gouyon, F., Schmidt, E. M., Celma, O., & Kim, Y. E. (2015). Modeling genre with the Music Genome project: Comparing human-labeled attributes and audio features. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 31–37). Malaga.

Raimond, Y., Abdallah, S., Sandler, M., & Giasson, F. (2007). The music ontology. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 417–422). Vienna.

Rhodes, W. (1965). The use of computer in the classification of folk tunes. *Studia Musicologica, VII*, 339–343.

Russell, J. A. (2003). Core affect and the psychological construction of emotion. *Psychological Review, 110*(1), 145–172.

Rzeszutek, T., Savage, P. E., & Brown, S. (2012). The structure of cross-cultural musical diversity. *Proceedings of the Royal Society B-Biological Sciences, 279*(1733), 1606–1612.

Savage, P. E. (2017). *Measuring the cultural evolution of music: With case studies of British-American and Japanese folk, art, and popular music* (PhD thesis). Tokyo University of the Arts, Tokyo.

Savage, P. E., & Brown, S. (2013). Toward a new comparative musicology. *Analytical Approaches to World Music, 2*(2), 148–197.

Savage, P. E., & Brown, S. (2014). Mapping music: Cluster analysis of song-type frequencies within and between cultures. *Ethnomusicology, 58*(1), 133–155.

Savage, P. E., Brown, S., Sakai, E., & Currie, T. E. (2015). Statistical universals reveal the structures and functions of human music. *Proceedings of the National Academy of Sciences of the United States of America, 112*(29), 8987–8992.

Savage, P. E., Matsumae, H., Oota, H., Stoneking, M., Currie, T. E., Tajima, A., ... Brown, S. (2015). How circumpolar is Ainu music? Musical and genetic perspectives on the history of the Japanese archipelago. *Ethnomusicology Forum, 24*(3), 443–467.

Savage, P. E., Merritt, E., Rzeszutek, T., & Brown, S. (2012). CantoCore: A new cross-cultural song classification scheme. *Analytical Approaches to World Music, 2*(1), 87–137.

Schaffrath, H. (1995). *The Essen folksong collection in the Humdrum Kern Format* (D. Huron, Ed.). Menlo Park, CA: Center for Computer Assisted Research in the Humanities.

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, E. G., & 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.

Scherrer, D. K., & Scherrer, P. H. (1971). An experiment in the computer measurement of melodic variation in folksong. *The Journal of American Folklore, 84*(332), 230–241.

Serrà, J., Corral, Á., Boguñá, M., Haro, M., & Arcos, J. L. (2012). Measuring the evolution of contemporary western popular music. *Scientific Reports, 2*(521). doi:10.1038/srep00521

Serra, X. (2011). A multicultural approach in music information research. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 151–156). Miami.

Serra, X. (2014). Creating research corpora for the computational study of music: The Case of the CompMusic Project. In *AES 53rd International Conference: Semantic Audio*. London.

Shalit, U., Weinshall, D., & Chechik, G. (2013). Modeling musical influence with topic models. In *Proceedings of the International Conference on Machine Learning* (pp. 244–252). Atlanta.

Shanahan, D., Neubarth, K., & Conklin, D. (2016). Mining musical traits of social functions in native American music. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 681–687). New York, NY.

Srinivasamurthy, A., Holzapfel, A., & Serra, X. (2014). In search of automatic rhythm analysis methods for Turkish and Indian art music. *Journal of New Music Research, 43*(1), 94–114.

Stevens, C. J. (2012). Music perception and cognition: A review of recent cross-cultural research. *Topics in Cognitive Science, 4*, 653–667. doi:10.1111/j.1756-8765.2012.01215.x

Temperley, D., & Van Handel, L. (2013). Introduction to the special issues on corpus methods. *Music Perception, 31*(1), 1–3.

Tenzer, M. (Ed.). (2006). *Analytical studies in world music*. New York, NY: Oxford University Press.

Thompson, D. (2015). 1991: The most important year in pop-music history. *The Atlantic*. Retrieved from http://www.theatlantic.com/entertainment/archive/2015/05/1991-the-most-important-year-in-music/392642/

Tzanetakis, G., & Cook, P. (2000). MARSYAS: A framework for audio analysis. *Organised Sound, 4*(3), 169–175.

Tzanetakis, G., Kapur, A., Schloss, A. W., & Wright, M. (2007). Computational ethnomusicology. *Journal of Interdisciplinary Music Studies, 1*(2), 1–24.

Underwood, T. (2015). *Can we date revolutions in the history of literature and music?*. Retrieved from https://tedunderwood.com/2015/10/03/can-we-date-revolutions-in-the-history-of-literature-and-music/

Underwood, T., Long, H., So, R. J., & Zhu, Y. (2016). *You say you found a revolution*. Retrieved from https://tedunderwood.com/2016/02/07/you-say-you-found-a-revolution/

Urbano, J., Bogdanov, D., Herrera, P., Gómez, E., & Serra, X. (2014). What is the effect of audio quality on the robustness of MFCCs and chroma features. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 573–578). Taipei.

van Kranenburg, P., Volk, A., & Wiering, F. (2013). A comparison between global and local features for computational classification of folk song melodies. *Journal of New Music Research, 42*(1), 1–18.

van Kranenburg, P., Garbers, J., Volk, A., Wiering, F., Grijp, L., & Veltkamp, R. C. (2010). Collaborative perspectives for folk song research and music information retrieval: The indispensable role of computational musicology. *Journal of Interdisciplinary Music Studies, 4*(1), 17–43.

Viro, V. (2011). Peachnote: Music score search and analysis platform. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 359–362). Miami.

Volk, A., & de Haas, W. B. (2013). A corpus-based study on ragtime syncopation. In *Proceedings of the International Society for Music Information Retrieval Conference* (pp. 163–168). Curitiba.

Volk, A., & van Kranenburg, P. (2012). Melodic similarity among folk songs: An annotation study on similarity-based categorization in music. *Musicae Scientiae, 16*(3), 317–339.

Volk, A., Wiering, F., & van Kranenburg, P. (2011). Unfolding the potential of computational musicology. In *Proceedings of the 13th International Conference on Informatics and Semiotics in Organisations (ICISO)* (pp. 137–144). Leeuwarden.

von Hornbostel, E. M., & Sachs, C. (1961). Classification of musical instruments. *Galpin Society Journal, 14*, 3–29.

Wallmark, Z. (2013). *Big data and musicology: New methods, new questions*. Retrieved from http://www.academia.edu/6442281/Big_Data_and_Musicology_New_Methods_New_Questions

Walshaw, C. (2014). A statistical analysis of the ABC music notation corpus: Exploring duplication. In *Proceedings of the Fourth International Workshop on Folk Music Analysis* (pp. 1–8). Istanbul. doi:10.13140/2.1.4340.0961

Wiering, F., & Benetos, E. (2013). Digital musicology and MIR: Papers, projects and challenges. In *ISMIR 2013 late-breaking session*. Retrieved from http://ismir2013.ismir.net/wp-content/uploads/2014/02/lbd4.pdf

Zhou, F., Claire, Q., & King, R. D. (2014). Predicting the geographical origin of music. In *IEEE International Conference on Data Mining* (pp. 1115–1120). Shenzhen.

Zivic, P. H. R., Shifres, F., & Cecchi, G. A. (2013). Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences of the United States of America, 110*(24), 10034-8.