The Interpersonal Entrainment in Music Performance Data Collection

MARTIN CLAYTON [1]
Durham University, United Kingdom

SIMONE TARSITANI
Durham University, United Kingdom

RICHARD JANKOWSKY
Tufts University, United States of America

LUIS JURE
Universidad de la República, Uruguay

LAURA LEANTE
Durham University, United Kingdom

RAINER POLAK
Max Planck Institute for Empirical Aesthetics, Germany

ADRIAN POOLE
Private Scholar

MARTÍN ROCAMORA
Universidad de la República, Uruguay

PAOLO ALBORNO
University of Genoa, Italy

ANTONIO CAMURRI
University of Genoa, Italy

TUOMAS EEROLA
Durham University, United Kingdom

NORI JACOBY
Max Planck Institute for Empirical Aesthetics, Germany

KELLY JAKUBOWSKI
Durham University, United Kingdom

ABSTRACT: The Interpersonal Entrainment in Music Performance Data Collection (IEMPDC) comprises six related corpora of music research materials: Cuban Son & Salsa (CSS), European String Quartet (ESQ), Malian Jembe (MJ), North Indian Raga (NIR), Tunisian Stambeli (TS), and Uruguayan Candombe (UC). The core data for each corpus comprises media files and computationally extracted event onset timing data. Annotation of metrical structure and code used in the preparation of the collection is also shared. The collection is unprecedented in size and level of detail and represents a significant new resource for empirical and computational research in music. In this article we introduce the main features of the data collection and the methods used in its preparation. Details of technical validation procedures and notes on data visualization are available as Appendices. We also contextualize the collection in relation to developments in Open Science and Open Data, discussing important distinctions between the two related concepts.

Submitted 2020 May 3; accepted 2020 September 10.
Published 2021 December 10; https://doi.org/10.18061/emr.v16i1.7555

KEYWORDS: *ethnomusicology, music cognition, computational musicology, entrainment, synchronization, music performance, Open Science, FAIR principles*

THE Interpersonal Entrainment in Music Performance (IEMP) project aimed to advance understanding of interpersonal entrainment – the synchronization and coordination of actions and resultant sounds between participants – in music performance (Clayton et al. 2005, Clayton et al. 2020). Empirical studies of musicians' behavior in natural performance situations are rare but important for a comprehensive understanding of music making, since laboratory experiments omit many contextual factors which may be highly relevant, while typically asking participants to produce or respond to highly simplified sequences. The project brought together various researchers and teams to share corpora, specialist knowledge and analytical approaches. The aim was to use comparative study to explore cultural variation in interpersonal entrainment. Partly for this reason, we decided to conduct analysis using existing audiovisual recordings that were obtained using commonly available video cameras and audio recorders, rather than relying on specialist set-ups such as Motion Capture facilities, thus minimizing the intrusiveness of the research. Analytical methods developed with these materials may also be adapted in the future to enable larger-scale corpus studies, including the use of video recordings obtained through web services such as YouTube.

The research questions, which included the variability of synchronization accuracy and precision and the role of body movement in ensemble coordination between diverse musical genres, required corpora including multitrack audio recordings (to facilitate event onset extraction for synchronization analysis) as well as video footage shot with static cameras (to facilitate the use of existing movement tracking algorithms for movement coordination analysis). No shared corpora meeting these criteria were available. This is a well-known problem for cross-cultural computational musicological analysis. Numerous other sources of audio and video recordings of diverse musical traditions exist, of course, from commercial recordings to curated audiovisual archives and items uploaded to web services. Most sources, however, lack reliable metadata and expert annotation; very little multitrack audio is freely shared, and high-quality video material using fixed cameras is also rare. Research collections are often retained by the researchers who recorded them and not shared with the wider community – sometimes for good reasons such as commercial or ritual sensitivities. The CompMusic project led by Xavier Serra addressed some of these issues in studying five diverse musical traditions (Serra, 2014). Corpora gathered for this project were compiled from existing recordings and are either freely shared or available to registered researchers and include extensive metadata: recordings are not multitrack however, and the collections do not include video.

In order to facilitate cross-cultural empirical and computational musical analysis, there is therefore an urgent need for more specialist corpora to be shared. The Interpersonal Entrainment in Music Performance Data Collection (IEMPDC) was therefore conceived by project team members as a contribution to the development of Open Science in music research, i.e., with replication of analyses and reuse of the data in mind. Beyond the immediate goals of the IEMP research project, by bringing these corpora together, annotating them to common standards and also sharing comprehensive documentation on their preparation – including code used for annotations such as onset extraction and assignment – we aim to create a resource of wide use in future empirical and computational music research. Such a benefit depends on a significant input of effort at the preparation stage. Standardizing the format of media files and checking their synchronization; standardizing the format of annotation files and fully documenting them; preparing versions of computer code to work reliably without external references; all of these tasks take time, and this time needs to be factored into project planning. In the case of IEMPDC, we also committed resources to technical validation: making independent annotations of metrical structure and comparing the results of onset detection and assignment achieved by different methods (the results can be found in Appendix A).

The collection described in this paper comprises audiovisual data and a range of annotations. Prominent amongst the annotations are two approaches central to our entrainment analyses: (a) sets of event onset data extracted from audio recordings and linked to metrical positions, and (b) movement data extracted from video. These can be reused directly for further study of entrainment or compared with annotations of the same features produced by different methods. Facilitating the use of onset and movement data are annotations of metrical structure and of musical form, and these annotations can be used as part of other kinds of analysis (e.g., of melody). The collection therefore creates possibilities for reuse well beyond the domain of interpersonal entrainment.

THE IEMP DATA COLLECTION

The six corpora making up IEMPDC are a subset of the potential datasets considered for entrainment analysis by IEMP. They meet our research criteria in terms of the availability of multitrack audio and, in four of six cases, static video shots, and were constructed from scratch (conceptualized, recorded, and annotated) by the researchers themselves in the context of their long-term engagement with the concerned styles of music and communities of musicians, spanning multiple projects often carried out over decades. IEMPDC therefore includes six diverse corpora, each with its own designated curator(s): those individuals responsible for identifying, selecting and annotating recordings – and in many cases also for recording them. Shared materials include audio and video recordings as well as annotations of musical meter and structural features, extracted event onsets and movement data extracted from the videos. Code used to prepare the corpora is also shared.

The reuse value of the recordings is as follows. First, our published analyses (Clayton et al. 2019, Clayton et al. 2020, Eerola et al. 2018, Jakubowski et al. 2017) can be replicated and adapted using alternative approaches, for example substituting different onset detection or movement tracking algorithms. Parameters considered to date could be combined with additional parameters (e.g. pitch, timbre) to enable further analysis of entrainment and other aspects of performance interaction as well as individual timing. Secondly, the availability of multitrack audio and video with expert annotations means that this collection can be used for a wide range of empirical investigations in computational musicology, music information retrieval, empirical musicology, and performance analysis: for example, in the study of gesture or of melodic, harmonic or phrase structure. Roeske et al. (2020) have used the collection to compare the rhythmic features of human music and birdsong. Thirdly, the approach taken here could be used as a model for similar analysis of new multitrack audio and/or video material, adding to the available corpora.

The following corpora make up the IEMP Data Collection:

- *Cuban Son & Salsa* (CSS, curated by Poole; Poole et al. 2019; Poole 2013)
- *European String Quartet* (ESQ, curated by Clayton; Clayton & Tarsitani 2019)
- *Malian Jembe* (MJ, curated by Polak; Polak et al. 2018; London et al. 2017, Neuhoff et al. 2017, Polak 2017, Polak 2020, Polak & London 2014, Polak et al. 2016)
- *North Indian Raga* (NIR, curated by Clayton and Leante; Clayton et al. 2018; Clayton et al. 2019)
- *Tunisian Stambeli* (TS, curated by Jankowsky; Jankowsky et al. 2019; Jankowsky 2010, 2013)
- *Uruguayan Candombe* (UC, curated by Jure and Rocamora; Jure et al. 2019; Fuentes et al. 2019, Jure & Rocamora 2018, Rocamora et al. 2015, Rocamora et al. 2019)
- *Technical Resources* (Eerola et al. 2019)

Other suitable sets of recordings were also identified and may form part of future expansions of this collection. Reference is also made to the "Improvising Duos" corpus (curated by Moran and Keller; Moran & Keller 2015, Clayton et al. 2017, Moran et al. 2017) which is analyzed for IEMP papers (Jakubowski et al. 2017; Eerola et al. 2018). This corpus was shared in a different format and on different repositories, however, and does not conform to the model described in this paper, therefore it is not covered here.

In order to carry out entrainment analyses and to maximize the usability of the collection for future researchers, the following steps were carried out. Preparation steps are summarized in Figure 1.

1. Identification of suitable recordings and preparation of appropriately matched media files.
2. Expert manual annotation of musical features such as metrical structures and formal features such as section boundaries. (Formal features annotated vary between corpora: for more information see the corpus documentation.)
3. Extraction of event onset data for all suitable instruments.
4. Assignment of selected event onsets to metrical positions, combining the data from 2 and 3.
5. Calculation of further derivative measures on this basis such as event density.
6. Extraction of movement data from selected video recordings, using computer vision algorithms.
7. Preparation of annotation data, code and corpus documentation for deposit in repositories.

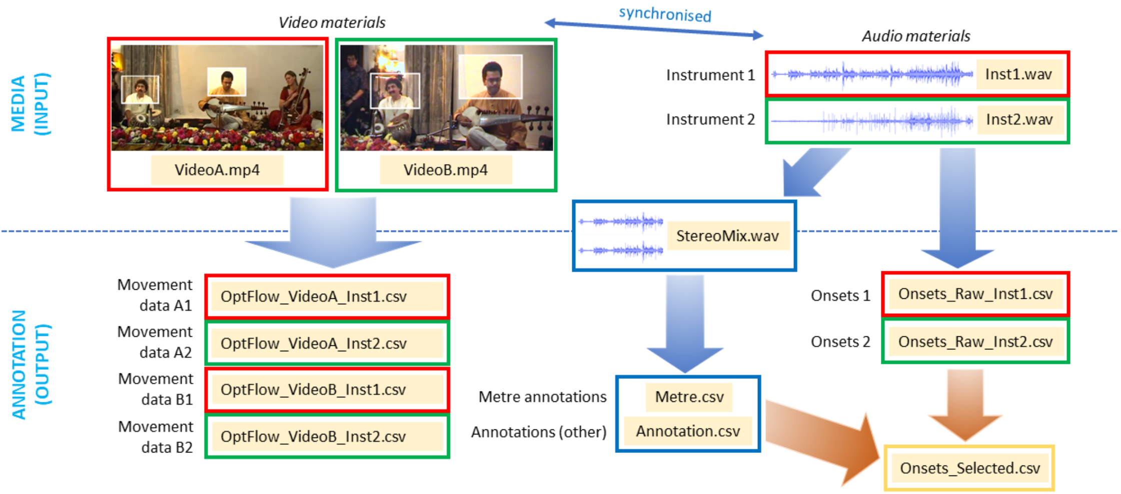


Fig. 1. Schematic representation of data preparation from original video (left) and audio (right) materials. File types are represented in beige-highlighted text.

Media Selection and Preparation

Details of the media for the IEMPDC corpora can be found in Table 1. As can be seen, the media originated from different sources and were recorded using a variety of formats and standards. While we aimed to apply consistent standards to the different corpora, therefore, this was not always strictly possible. Our procedure in preparing the collection was the following:

1. Select discrete items (pieces, takes, or sections). Selection criteria included the existence of good quality audio with separation between parts, and where possible good quality video taken with static camera angles, consistent lighting and minimal occlusion. (Assurance of musical performance quality was covered in the selection of materials offered by researchers for the project and was not a significant factor in choosing between pieces at this stage.)

2. Save linked and matched media files to standards and formats common across the collection.

Depending on how the collection was originally shared, this process could require either synchronization of media (e.g. multitrack audio and multiple video files), or simply preparing the files as needed for sharing (e.g. transcoding, renaming, etc.).[2] Multiple files of each single item have the same in-out points (and therefore also the same duration). Audio files were saved as mono or stereo WAV files, keeping the bitrate and sampling frequency they were recorded and/or edited in. Video files were transcoded as MP4 files, using H264 MPEG-4 AVC (part 10) codec; the transcoding was processed on Squared 5 “MPEG Streamclip” software, setting the encoding quality to 60% without setting a limit on the data rate. SD videos were kept in their original resolution, while Full HD videos were scaled down to 720p HD (1280x720). The video formats have been chosen as a balanced compromise between quality and usability of the files on the online sharing platform. The following sections describe the procedures used to produce output data. Annotation data are summarized in Table 2.

Table 1. *Summary of Raw Data (Media Files) by Sub-Project*

Corpus	Content	Duration	Audio Files	Video Files
CSS	5 songs recorded by the 7-piece Cuban band Asere, in son and salsa styles.	c. 33 mins	Multi-track audio. Up to 15 mono tracks per song plus stereo mix (WAV, 48 kHz, 24 bit). 75 files.	3 views recorded as PTZ camera direct to hard disk recording, MOV, PAL, SD, converted to DVD (MPEG-2 720x576 25 fps), exported as MP4, H264 MPEG-4 AVC (part 10) with stereo mix of audio substituted for camera sound. 15 files.
ESQ	2 takes each of two string quartet movements: the first movement of Haydn's Op. 76 No. 5, and the third of Beethoven's Op.59 No.2.	c. 20 mins	Multi-track audio. 4 mono tracks plus stereo mix per take (WAV, 48 kHz, 24 bit). 20 files.	None.
MJ	15 recordings of 3 repertoire items of Malian jembe ensemble music including drum duos, trios, and quartets.	c. 50 mins	Multi-track audio using clip-on microphones (AKG C-417) clipped to the drum-rims, very close to membrane; recorded into Edirol-R4 mobile four-track studio. 2-4 mono tracks per take (WAV, 48 kHz, 16 bit). 46 files.	1 view per track recorded as AVI from mini-DV, 720x560, 25 fps progressive scan, exported as MP4, H264 MPEG-4 AVC (part 10) with stereo mix of audio substituted for camera sound. 15 files.
NIR	8 complete raga performances comprising 4 instrumental gats (for 2 of which media are split into 2 sections each); 3 khyal vocal; and 1 tabla solo.	c. 414 mins	Multi-track audio. Mono for each instrument plus stereo mix for each piece (WAV, 44.1 or 48 kHz, 16 or 24 bit). 49 files	2-3 views of each performance shot in either SD (4:3) or HD (16:9) format, encoded as MP4, H264 MPEG-4 AVC (part 10) with stereo mix of audio substituted for camera sound. 24 files.
TS	4 pieces of Tunisian stambeli ritual music, featuring gumbri (lute), shqashiq (cymbals) and voices; recorded for a commercial CD release.	c. 40 mins	Multi-track audio. Up to 8 mono tracks per song (WAV, 48 kHz, 24 bit). 28 files.	None.
UC	12 takes of Uruguayan candombe music, including drum trios and quartets.	c. 35 mins	Multi-track audio. Up to 4 mono tracks per song plus stereo mix (recorded as WAV, 48 kHz, 24 bit, output as WAV 44.1kHz, 16 bit). 51 files.	1 view per take recorded as MOV, H264 MPEG-4 AVC, 1920x1080, 43959 kb/s, 23.98 fps, exported as MP4, H264 MPEG-4 AVC (part 10) with stereo mix of audio substituted for camera sound. 12 files.

Note. CSS = Cuban Son & Salsa, ESQ = European String Quartet, MJ = Malian Jembe, NIR = North Indian Raga, TS = Tunisian Stambeli, UC = Uruguayan Candombe

Table 2. *Annotation Data.*

Corpus	Onsets_Raw	Onsets_Selected	Movement data	Metre annotations	Annotations (other)
CSS	One combined CSV file per song with all instruments combined. 5 files.	One file per song, including onset times, peak levels and density calculations plus clave pattern and section indications (son/montuno). 5 CSV files.	Optical flow data for heads and feet of two musicians for three songs, plus 5 files specifying regions of interest (ROI). 12 CSV, 5 TXT files.	One file per song. 5 CSV files.	Music structure and interaction between two musicians. 5 CSV files. Documentation also includes song texts.
ESQ	One combined CSV file per piece with all instruments combined. 4 files.	One combined file per piece, including onset times, peak levels and density calculations for staccato portions only (Haydn bars 41-46 and 76-127, Beethoven bars 132-169). Annotation of lead instrument in Haydn. 4 CSV files.	None.	One file per take. 4 CSV files.	No separate files.
MJ	One CSV file per instrument per track. 46 files.	Original set: One CSV file per take, including onset times (Polak/Jacoby) and density calculations. 15 files. New set: One file per take, including onset times, peak levels and density calculations. 15 CSV files.	Optical flow data for heads of all musicians for each of three takes plus one ROI file. 9 CSV, 1 TXT files.	One file per take. 15 CSV files.	Song form (sections and signal patterns) annotated for all takes. 15 CSV files.
NIR	For instrumental gat pieces, one CSV file per instrument (solo + tabla). 10 files. Not included for vocal or tabla solo performances.	For instrumental gat pieces, one file per tala (meter) section. Includes metrical position [= matra (beat), half-matra for slow sections]; onset times, peak levels and density calculations; cadential downbeats; segmentation used in published analyses. Not included for vocal or tabla solo performances. 10 CSV files.	Optical flow data for all sections. CSV files include x and y coordinates of barycenter of Regions of Interest selected around head and upper torso of musicians selected on video files plus 8 ROI files. 23 CSV, 8 TXT files.	One file per tala (metre) section. 18 CSV files.	For all: Start-End and Form (timings of section breaks). Depending on piece: Interaction; Gesture; Notes (misc.). 10 CSV files. Documentation also includes song texts.
TS	One CSV file per instrument per take. 9 files.	1 file per piece, including onset times for each instrument, peak levels and density calculations, stroke and beat numbers, sections and segmentation. 8 CSV files.	None.	None.	None.
UC	One CSV file per instrument per take. 39 files.	One combined file per take, including onset times, peaks, density calculations and repique drum parts (quartets only). 5 CSV files.	None.	One file per take. 12 CSV files.	No separate files.

Note. CSS = Cuban Son & Salsa, ESQ = European String Quartet, MJ = Malian Jembe, NIR = North Indian Raga, TS = Tunisian Stambeli, UC = Uruguayan Candombe

Onset Extraction (Onsets_Raw)

Event onsets were extracted for all suitable instruments across the collection. Suitable instruments are those with attacks (envelope onsets) sufficiently steep that they would be reliably found by existing computational algorithms. Excluded sounds include the voice (CSS, NIR, TS), tanpura (accompanying lute, NIR), and shakers (CSS); for the bowed strings in the ESQ corpus staccato sections only were used. Event onsets were extracted using a specially created algorithm which output both onset time and peak level data (see Eerola et al. 2019). Onsets had been independently extracted for both MJ and UC corpora prior to assembly of the collection (see Appendix A for more details): the final versions of the onset data nonetheless substitute those extracted using the new algorithm to ensure consistency. (In the MJ corpus, both sets of assigned onsets are included in the published corpus.)

Onsets were extracted based on envelope characteristics using MIR Toolbox (Lartillot et al. 2007). First the audio signal is band-pass filtered to focus on the most relevant frequency band. The envelope of the filtered signal is then extracted and subjected to low-pass filtering and half-wave rectification before applying peak-picking algorithms with three parameters. These parameters determine (1) the local contrast threshold value, (2) normalized amplitude threshold value, and (3) threshold value for the minimum difference between peak values. The rationale for the reliance on one particular onset extraction technique that relies on envelope energy instead of other possible components such as spectral flux, for instance, was to keep the onset timings comparable across instruments in terms of the lag inherent in onset detection. One advantage of this method is that unlike most commonly available onset detection programs it does not quantize results (this is typically done to intervals of 10 ms or larger) but uses interpolation to achieve greater temporal precision.

No onset detection algorithm perfectly matches human judgements of the occurrence of events (perceptual onsets). Any audio event onset time is an estimate, either of the physical onset of the signal (as in IEMPDC) or of the perceived onset (p-center): as with meter annotations (see below), time points are estimates. For some instruments, for example drums playing clear patterns with little variation in timbre and dynamics, the margin for error is very small (identifying the occurrence of an onset is uncontroversial and the time can be estimated within a few milliseconds). For others, such as the sitar (NIR), the variety of onset types and the wide dynamic range means that if a single set of parameters is applied across a whole recording, then a compromise must be reached. Our aim in setting extraction parameters was to ensure that the vast majority of humanly-identified onsets were captured accurately, while generating a relatively low number of false positives (see Appendix A). A small number of onsets could not be extracted using this approach (ca. 5%, see Appendix A): in many cases they could have been marked manually, but we have not taken this approach to avoid mixing onsets derived by means of a known algorithm with manually identified onsets.

Parameter setting followed different procedures for different instruments (procedures and parameters are reported in the corpus documentation files on Open Science Framework (OSF; https://osf.io/37fws/)). For most instruments, especially drums and other percussion, excellent extraction was possible by (a) setting the frequency parameters based on visual inspection of spectrograms, and then (b) adjusting sensitivity settings by trial and error. In some cases, parameters were adjusted within corpora between items (takes). Some instruments proved too challenging to deal with in this way, especially the Indian instruments, whose sounds are spectrally complex and have a very wide dynamic range. In this case parameters were optimized automatically by preparing manually annotated 1-minute samples to act as ground truth against which the algorithm performance was compared using an F-measure and 70 ms tolerance for the correct onsets, reaching F-measures between 0.75 and 0.95; files and scripts used are shared as part of the collection.

Since recordings were all made in live ensemble performances with conventional (not contact) microphones, some bleed is found between tracks (i.e. the sound of one instrument can be heard on another instrument's audio track). In most cases this is not a significant problem, since onset detection parameters can be tuned to exclude unwanted onsets (e.g. by setting an appropriate frequency range). In a few cases this was a more challenging problem. For example, conga and bongo drum onsets are easily confused (CSS), and significant manual intervention was required to minimize this problem.

Meter Annotations

Each recorded section was manually annotated by an expert in the particular musical style. For all metrical sections – comprising the great majority of the collection, the main exception being introductory alap sections in NIR, which have no regular beat – the metrical boundaries (and for longer cycles, intermediate points) were manually tapped and recorded in Sonic Visualiser. "Metrical sections" for this purpose are those featuring a clear beat and its organization into repeating patterns of fixed numbers of beats; annotated metrical boundaries are the main downbeats. For TS, whose songs are based on fixed repeated rhythmic patterns, the shqashiq (cymbal) rhythms marking out repeated rhythmic patterns were initially subject to manual onset identification in Sonic Visualiser. Where metrical errors such as dropped beats occur, these were annotated (this occurs rarely in NIR, not in the other corpora). In the represented musical traditions, identifying metrical boundaries is not controversial and their position can be estimated by manual tapping by a listener familiar with the style (see Appendix A). Where tapping errors were identified, for example because the manually tapped time was sufficiently inaccurate to adversely affect the onset assignment process (see below), metrical annotations were subsequently manually corrected. It should be noted that although these manual annotations guided the assignment of event onset times to metrical positions, they do not affect the extracted onset data. They are sufficiently accurate to enable windows to be defined within which to search for event onsets, and to inform visual representations (see Figure B2 in Appendix B).

Onset Assignment and Manual Checking (Onsets_Selected)

The aim in this step was to allocate extracted onset times to specific metrical positions and then check them for accuracy. This process was carried out using the same method for all corpora, using Excel spreadsheets. It was also carried out independently using a different approach for MJ and UC, and comparison between the results was used for technical validation (see Appendix A). The common approach is described first.

Nominal metrical positions were identified for each corpus. In most cases the manually-tapped metrical cycles were simply divided into a number of equal beat durations. In the slowest sections of NIR, where the nominal beat (matra) can be over 1 sec long, half-beats were also identified at this stage. In MJ two of the three repertory items use a non-isochronous ternary subdivision of four main beats to generate 12 metrical positions, and in this case the division of the cycle used the mean subdivision proportions reported in published analyses (Polak, 2010; Polak et al., 2016). For TS, since the rhythm is based on repeated stroke patterns which can be metrically ambiguous (Jankowsky, 2010, 2013), the main strokes of each pattern were identified directly.

The resulting time points are described as “Virtual beats”: they do not form part of the final output data but are used to identify windows within which to search for onsets. The initial assignment of onsets to metrical positions was achieved by selecting the closest event onset to each Virtual beat if it was found within a specified window. Windows were set for each recorded item, and where necessary adjusted within items (e.g. in the case of significant tempo change) in such a way that the great majority of onsets are captured (>95%) but incorrect assignments are minimized. The size of selection windows varied between +/-30 ms and +/-160 ms, depending on the beat subdivision: for example, with a beat subdivision as short as 100 ms the window needs to be set below 50 ms to avoid onsets being assigned to the wrong metrical position, while in much slower passages in the NIR corpus an onset might occasionally fall as much as 150 ms off a predicted position and still be perceived as falling “on” that position (i.e. as belonging to a particular position but played early or late for expressive purposes, rather than belonging to another metrical position). Window sizes are reported in more detail in corpus documentation files on OSF.

Onset assignment for NIR followed a slightly more complicated procedure due to fluctuations of the beat duration within cycles. In these pieces tabla onsets were assigned first, since speed can fluctuate within a metrical cycle and identifying the appropriate tabla onsets is generally the simplest way of estimating each beat position. Melody instrument onsets were identified first in a window around the tabla onsets; where no tabla onset had been identified, they were identified within windows around the virtual beats. Onsets assigned in this way were then visualized as labels within Sonic Visualiser (Cannam et al. 2010) and manually checked. The following were manually removed:

i. false positives (i.e. the onset was not recognized as such by the listener, or did not match the humanly-perceived onset), and

ii. wrongly-assigned onsets (i.e. although an onset was correctly identified and it fell within the window, an expert judgement was made that it was not intended to be heard on the specific metrical position).

Relatively simple cases included triplet variations, in which a stroke falling halfway between two metrical positions fell within the selection window and had to be manually removed. In the more complicated examples, especially in NIR, tempo fluctuated enough within a cycle to cause significant windowing errors;

in the same corpus the complexity of some rhythmic variations means that analytical judgements had to be made as to what rhythmic relationship was being performed before onsets could be meaningfully assigned. Where appropriate, onsets were also manually reassigned to their correct metrical positions; the onset times themselves were not adjusted in any way, however.

As noted above in the case of MJ and UC, the processes of windowing and manually checking onset assignment had been completed independently by the curators. The assignment process used for MJ and UC was as follows: the instrument playing the most regular ostinato pattern (e.g Jembe 2 or Dundun 1 in MJ, Chico in UC) was used to identify cycle beginnings. All other instruments that articulated each downbeat (using a window of +/- 100 ms compared with the ostinato instrument) were then identified, and the metrical boundary was defined as an average of the onsets of all instruments articulating the downbeat. For each onset in the database, the relative location within the cycle was calculated and histograms computed for these relative positions, identifying one peak for each of the subdivision positions within the cycle. Mean peak locations were computed for each metrical position and a window calculated from the normalized beat duration was defined. In the case of MJ, the window extended to 17% of the normalized beat duration for each of the three subdivisions spread asymmetrically (−10% to +7%) around the mean value for each, and a small percentage of all onsets outside that window were discarded (the asymmetry was in order not to erroneously discard strokes of Jembe 1, which tends to play slightly earlier than the other drums). For UC, there are four subdivisions per beat and the window was symmetrical (+/-12%). In cases with multiple onsets in each bin, the onset that was closest to the bin center was selected. This was done in order to remove onsets composing part of an ornamental figure such as a flam. The total number of removed onsets constituted less than 3% of the entire database. The Matlab script used for this process is shared with IEMPDC (see Onset assignment (Matlab) in Eerola et al. 2019).

The final Collection output data (Onsets_Selected files) for MJ and UC is based on the curators' onset assignments derived as in the previous paragraph, with minor adjustments (e.g., the script omitted the first and/or last cycles of takes in MJ). The onset times themselves, however, were substituted with those produced by the common detection algorithm (see above), by searching for the closest onset within a window of +/- 40ms of the curators' estimate: >99% of onsets were successfully matched this way.

For TS, curator Richard Jankowsky had manually marked onsets for the shqashiq (cymbals) part for each of the main rhythmic strokes. These were replaced with onsets produced by the common script and falling within a window of +/-70ms, and then gumbri (lute) onsets matched within a fixed window of the shqashiq onsets (65-80 ms, depending on the piece), before all onset assignments were manually checked. The resulting Outputs_Selected files therefore result from a combination of purely algorithmic selection and expert human judgement. The files include columns for metrical position and onset time for each instrument, with corresponding peak levels. Calculations of local event density are also included (extracted onsets per second over the preceding 2 seconds): this is a derived measure and could be easily recalculated, nonetheless the figures used for published papers are shared.

Movement extraction

As part of the IEMP project, computer vision algorithms were implemented in EyesWeb XMI 5.6.2.0 (http://www.infomus.org/eyesweb_eng.php). Movement extraction output employing the Optical Flow algorithm (Jakubowski et al. 2017) is shared for specific parts of the collection. These include all of NIR (main performers only, i.e., soloists, tabla and harmonium accompanists), 3 video files from MJ, and 3 songs from CSS (two musicians only).

Optical flow (OF) is an established technique within the computer vision literature, based on calculation of the distribution of apparent velocities of movement of brightness patterns in an image (Farnebäck 2003). The OF algorithm outputs Cartesian coordinates for a calculated barycenter of the moving mass within a designated Region of Interest (ROI). ROIs were manually selected around body parts of interest, typically heads and upper torsos (CSS also includes some data for feet). ROI data and the output for each ROI used (the barycenter coordinates) is shared as part of the collection, as is the OF patch for the freely-shared EyesWeb program (Windows; see Eerola et al., 2019).

Table 3. Technical Resources shared as part of IEMPDC (Eerola et al., 2019).

Component name	**Usage**	**Details**
Onset detection	Onset detection (all corpora)	Matlab script to extract onset times and peak levels, plus scripts and files for validation of parameters. Used for all corpora.
Onset assignment (Excel)	Onset assignment (CSS, NIR, ESQ, TS)	Excel spreadsheet for assignment of onsets and calculation of event density.
Onset assignment (Matlab)	Onset assignment (MJ, UC)	Matlab script to establish metrical framework and assign onset times to metrical positions. Used for MJ and UC corpora.
Video tools	Video movement extraction	EyesWeb patches created for IEMP: only the Optical Flow patch was used in preparation of this collection.

Additional annotations

The corpora also include a range of other forms of annotation. Other columns in Onsets_Selected files, for example, include performers' names, formal sections, manual segmentation of long items into arbitrarily-defined sections (used for analysis of NIR and TS; Clayton et al., 2019, Clayton et al. 2020) and identification of cadential downbeats (NIR).

Formal features included in Onsets_Selected annotations vary between corpora. In CSS the main division is between son and montuno sections, in MJ theme and variation sections are distinguished, tabla solo passages are marked in NIR, and vocal and instrumental sections are marked in TS. These annotations could be augmented according to the needs of specific analyses.

Selected recordings in CSS and NIR include annotations of visible bouts of interaction between performers. Bouts of interaction are defined here as "periods of interaction arising from the behavior of the performers, where the characteristic movement patterns of the two musicians indicated a degree of correspondence in the eyes of the annotator" (Eerola et al., 2018, p. 5). These annotations were all made in the annotation tool ELAN (https://archive.mpi.nl/tla/elan; Sloetjes & Wittenburg, 2008) and shared as CSV files. Technical resources enabling replication of annotations are summarized in Table 3.

Ethics, Permissions and Licenses

All musicians recorded as part of the collection gave their informed consent for the recordings to be shared. Specific procedures and approvals vary between corpora, as each recordist worked within their own institutional norms.

- NIR and ESQ are shared under a Durham University license (see https://osf.io/j8adp/);
- CSS, MJ, TS, and UC and Technical Resources under Creative Commons CC-BY-NC-ND 4.0 (see https://creativecommons.org/licenses/by-nc-nd/4.0/). The aim of both licenses is the same, which is to offer free access to research users, while restricting misuse and all commercial applications.

The following uses of the Collection are **not** permitted:

- any commercial exploitation
- unauthorized creative reuse, e.g. sampling or remixing, even where not for commercial gain
- unauthorized resharing on other public platforms (e.g. YouTube, SoundCloud)

Researchers should contact corpus curators in case of any doubt about the appropriateness of their intended usage.

FAIR PRINCIPLES, OPEN DATA AND FUTURE DIRECTIONS

As noted above, our aim in sharing these corpora was to make a significant contribution to Open Science in musicological research by making available high quality annotated audiovisual materials that can be easily used for a variety of empirical studies. We have found in practice that sharing and publicizing the collection was not as simple as we had hoped, and this experience prompts some reflection on the term "open data" and the FAIR principles (Wilkinson et al. 2016). Our initial ambition was in fact to publish an account of the collection in a general scientific journal specializing in data collections. Our efforts to potentially reach a wider readership through this demonstration of scientific rigor did not succeed, however, since our target journal declined to review the submitted article on the grounds that it did not adhere to their open data policy. It is not difficult to see why in retrospect: according to the Open Data Handbook, "Open data is data that can be freely used, re-used and redistributed by anyone - subject only, at most, to the requirement to attribute and sharealike" (https://opendatahandbook.org/guide/en/what-is-open-data/, accessed 17th Jan 2020). The European Data Portal suggests, "Open data must be licensed. Its licence must permit people to use the data in any way they want, including transforming, combining and sharing it with others, even commercially." (https://www.europeandataportal.eu/elearning/en/module1/#/id/co-01, accessed 20th January 2020).

Such an open data policy is not possible with the IEMPDC audiovisual recordings, since they comprise musical performances by expert musicians. Apart from the Tunisian Stambeli musicians, who recorded for a commercial release for which they were recompensed, the musicians in IEMPDC recorded their performances having given informed consent for their research use. If we were to share media on open data terms, there would be nothing to stop a record producer downloading the media, editing or remixing it and releasing the results publicly in ways that might (a) generate a financial gain which would not benefit the performers, and/or (b) present their music in a way that the performers would find objectionable or offensive. Previous examples of the appropriation of field recordings for commercial gain are well known within ethnomusicology (e.g., Feld, 2000; Tan, 2008; Zemp, 1996). While we cannot physically prevent such an occurrence, it is our duty to clearly point out the license terms relating to the shared recordings. Our view is that by releasing the data for unrestricted download under a license that stipulates that uses other than research are not permitted, we achieve the aims of open science while also maintaining appropriate ethical standards. We should not forget, after all, that ethical standards are an integral part of good science, and in the case of music research, respecting and protecting the interests of participants (including performers we record) is an essential element of good ethical practice. The fact that a high-profile scientific journal was not willing to consider this essential factor but chose to insist on a rigid definition of open data creates difficulties for music research: it suggests the existence of barriers to the implementation of computational (ethno)musicological research as a form of open science, since such an approach cannot be reconciled with ethical practice.

It is important to note at this point that completely open access to data is *not* a requirement of the FAIR principles of Open Science. The FAIR principles set out criteria for best practice in data sharing, covering Findability, Accessibility, Interoperability and Reusability. FAIR's paragraph explaining criterion A1.2 ("The protocol allows for an authentication and authorisation where necessary") is worth quoting here:

> *This is a key, but often misunderstood, element of FAIR. The 'A' in FAIR does not necessarily mean 'open' or 'free'. Rather, it implies that one should provide the exact conditions under which the data are accessible. Hence, even heavily protected and private data can be FAIR.* (https://www.go-fair.org/fair-principles/a1-2-protocol-allows-authentication-authorisation-required/, accessed 18th January 2020)

It appears then that FAIR principles and Open Data definitions are by no means aligned, and prominent journals are choosing to favor the latter over the former. One suggestion we would make is that Open Science gatekeepers should implement the former rather than the latter. The FAIR principles offer a useful framework for future music corpus development, and the additional burden of implementing them is surely justified by the scientific benefits.

The advantages of archiving IEMPDC on the Open Science Framework (OSF) include a user-friendly interface that allows intuitive navigation, flexible options for downloading data and excellent previewing of media files. OSF is thus optimized for the human user. In terms of machine readability, which is a key element of the FAIR principles, more work is required to facilitate the download of files and metadata, which is in principle possible using OSF's API and packages for R (https://github.com/ropensci/osfr) and Python (https://github.com/osfclient/osfclient). At the time of writing, we have not fully exploited the possibilities that this affords for streamlining analysis and visualizations. Possibilities for future access to the data and code include, for example, Jupyter notebooks (https://jupyter.org/), which allow the embedding of fragments of computer code within web pages and thus allow users to run analyses without installing specialist software. Such additions should be regarded as complements to the data collections themselves but offer advantages in promoting access and usability of IEMPDC and similar resources. Future corpora for scientific music research should aim to comply with FAIR principles alongside other important frameworks (including appropriate ethical standards). Further discussion of appropriate repositories, data formats and other important considerations is to be welcomed, and hopefully in some respects IEMPDC offers a useful model to the community.

AUTHOR CONTRIBUTIONS

PI & Lead author:

M.C. PI of IEMP project, curator of NIR and ESQ. Onset extraction, assignment and manual checking, manual annotation and documentation, technical verification. Drafting of text and figures.

Lead Technician:

S.T. Research technician in charge of media preparation. Recordist of CSS, ESQ and part of NIR.

Curators (in alphabetical order):

R.J. Curator of TS; manual annotations and documentation.
L.J. Co-curator of UC; recording, onset extraction, manual annotations, onset assignment and documentation.
L.L. Co-curator of NIR corpus, technical verification, Figure 1.
R.P. Curator of MJ; recording, onset extraction, manual annotations and documentation.
A.P. Curator of CSS; manual annotations and documentation.
M.R. Co-curator of UC; recording, onset extraction, manual annotations, onset assignment and documentation.

Data enrichers (contributed software tools, data extraction and organization; in alphabetical order):

P.A. Author of EyesWeb patches.
A.C. Co-PI. EyesWeb patches.
T.E. Co-PI. Matlab onset extraction script, parameter optimization, curator of technical resources.
N.J. Matlab windowing script; onset assignment (MJ).
K.J. Video movement extraction, text editing

ACKNOWLEDGMENTS

The Interpersonal Entrainment in Music Performance project was funded by the Arts and Humanities Research Council (grant no. AH/N00308X/1). Parts of the collection were funded additionally by:

- *Cuban Son & Salsa Corpus (CSS)*: Open University
- *European String Quartet Corpus (ESQ)*: BBC Radio 3
- *Malian Jembe (MJ):* Deutsche Forschungsgemeinschaft (DFG), research grant PO627/3-1, grant holder: Rainer Polak.
- *North Indian Raga (NIR):* AHRC grant no. MRG-AN6186; British Academy/Leverhulme Trust, SRG 2015-16 Round; St Chad's College Durham, Durham University and Open University.
- *Tunisian Stambeli (TS):* Par les Chemins Productions (France).
- *Uruguayan Candombe (UC):* Comisión Sectorial de Investigación Científica (CSIC) and Agencia Nacional de Investigación e Innovación (ANII), Uruguay.

The authors also wish to thank other members of the IEMP project team for their suggestions and support, all recordists and annotators involved in data preparation, and particularly the musicians who gave permission for their performances to be shared in this way. This article was copyedited by Jessica Pinkham and layout edited by Diana Kayser.

NOTES

[1] Correspondence can be addressed to: Professor Martin Clayton, Department of Music, Durham University, Palace Green, Durham, DH1 3RL, U.K. martin.clayton@durham.ac.uk.

[2] Different corpora (and in the case of NIR, different pieces within a corpus) were recorded using different setups. Synchronization between audio and video media was in some cases ensured using a common timecode generator (and in some cases genlock), while in other cases audio and video were recorded separately and synchronized at the editing stage. When manually synchronizing media, the accuracy is determined by the video frame rate: the synchronization error is usually within half of the duration of one video frame (for video files recorded at 25fps, one frame lasts 40ms and the synchronization error is therefore estimated to be lower than 20ms). Depending on the recording equipment time drift can occur, especially with long recordings (e.g. more than 30 minutes): when such issues have been noticed, they have been corrected.

[3] This is possible for the Maraka pieces since the subdivision is roughly isochronous; other MJ repertory items use non-isochronous subdivisions.

REFERENCES

Cannam, C., Landone, L. & Sandler, M. (2010). Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the ACM Multimedia 2010 International Conference*. https://doi.org/10.1145/1873951.1874248

Clayton, M., Eerola, T., Jakubowski, K. & Tarsitani, S. (2017). *Interactions in Duo Improvisations*. [Data Collection]. Colchester, Essex: UK Data Archive. https://doi.org/10.5255/UKDA-SN-852847

Clayton, M., Jakubowski, K. & Eerola, T. (2019). Interpersonal entrainment in Indian instrumental music performance: Synchronization and movement coordination relate to tempo, dynamics, metrical and cadential structure. *Musicae Scientiae*, *23*(3), 304–331. https://doi.org/10.1177/1029864919844809

Clayton, M., Jakubowski, K., Eerola, T., Keller, P., Camurri, A., Volpe, G. & Alborno, P. (2020). Interpersonal entrainment in music performance: Theory, method and model. *Music Perception*, *38*(2), 136–194. https://doi.org/10.1525/mp.2020.38.2.136

Clayton, M., Leante, L., & Tarsitani, S. (2018). *IEMP North Indian Raga*. https://doi.org/10.17605/OSF.IO/KS325

Clayton, M., Sager, R. & Will, U. (2005). In time with the music: The concept of entrainment and its significance for ethnomusicology. *European Meetings in Ethnomusicology 11* (ESEM Counterpoint 1), 1–82. http://dro.dur.ac.uk/8713/

Clayton, M., & Tarsitani, S. (2019). *IEMP European String Quartet*. https://doi.org/10.17605/OSF.IO/USFX3

Danielsen, A. (2018). Pulse as dynamic attending: Analysing beat bin metre in neo soul grooves. In C. Scotto, K. M. Smith & J. Brackett (Eds.), *The Routledge Companion to Popular Music Analysis: Expanding Approaches*. Routledge ISBN 9781138683112. https://doi.org/10.4324/9781315544700-12

Eerola, T., Clayton, M., Alborno, P., Camurri, A., Jacoby, N., Jakubowski, K., & Tarsitani, S. (2019). *IEMP Technical Resources*. https://doi.org/10.17605/OSF.IO/NVR73

Eerola, T., Jakubowski, K., Moran, N., Keller, P. E. & Clayton, M. (2018). Shared periodic performer movements coordinate interactions in duo improvisations. *Royal Society Open Science*, *5*(2). 171520. https://doi.org/10.1098/rsos.171520

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In J. Bigun & T. Gustavsson (Eds.), *Image Analysis. SCIA 2003. Lecture Notes in Computer Science*, vol. 2749. Springer: Berlin, Heidelberg. https://doi.org/10.1007/3-540-45103-X_50

Feld, S. (2000). A sweet lullaby for world music. *Public Culture*, *12*, 145–171. https://doi.org/10.1215/08992363-12-1-145

Fuentes, M., Maia, L. S., Rocamora, M., Biscainho, L. W. P., Crayencour, H. C., Essid, S. & Bello, J. P. (2019). Tracking beats and microtiming in Afro-Latin American music using conditional random fields and deep learning. *20th Conference of the International Society for Music Information Retrieva*l, Delft, Netherlands, 4-8 Nov 2019 (pp. 1–8). https://iie.fing.edu.uy/publicaciones/2019/FMRBCEB19/

Jakubowski, K., Eerola, T., Alborno, P., Volpe, G., Camurri, A. & Clayton, M. (2017). Extracting coarse body movements from video in music performance: A comparison of automated computer vision techniques with motion capture data. *Frontiers in Digital Humanities: Digital Musicology*, *4*, 9. https://doi.org/10.3389/fdigh.2017.00009

Jankowsky, R. (2010). *Stambeli: Music, Trance, and Alterity in Tunisia*. Chicago: University of Chicago Press. https://doi.org/10.7208/chicago/9780226392202.001.0001

Jankowsky, R. (2013). Rhythmic elasticity and metric transformation in Tunisian stambeli. *Analytical Approaches to World Music*, *3*(1), 34–61. http://www.aawmjournal.com/articles/2014a/Jankowsky_AAWM_Vol_3_1.pdf

Jankowsky, R., Tarsitani, S., & Clayton, M. (2019). *IEMP Tunisian stambeli*. https://doi.org/10.17605/OSF.IO/SWBY6

Jure, L. & Rocamora, M. (2018). Subir la llamada: negotiating tempo and dynamics in Afro-Uruguayan candombe drumming. *Proceedings of the 8th International Workshop on Folk Music Analysis (FMA),* Thessaloniki (pp. 25–30). http://fma2018.mus.auth.gr/files/papers/FMA2018_paper_6.pdf

Jure, L., Rocamora, M., Tarsitani, S., & Clayton, M. (2019). *IEMP Uruguayan Candombe*. http://fma2018.mus.auth.gr/files/papers/FMA2018_paper_6.pdf

Lartillot, O. & Toiviainen, P. & Eerola, T. (2007). A Matlab toolbox for music information retrieval. *Data Analysis, Machine Learning and Applications*, *4*, 261–268. https://doi.org/10.1007/978-3-540-78246-9_31

London, J., Polak, R. & and Jacoby, N. (2017). Rhythm histograms and musical meter: A corpus study of Malian percussion music. *Psychonomic Bulletin & Review*, *24*(2), 474–80. https://doi.org/10.3758/s13423-016-1093-7

Moran, N. & Keller, P. E. (2015). *Improvising Duos*, [moving image]. University of Edinburgh. https://doi.org/10.7488/ds/251

Moran, N., Hadley, L. Bader, M. & Keller, P. E. (2015). Perception of 'back-channeling' nonverbal feedback in musical duo improvisation, *PLoS ONE*, *10*(6). https://doi.org/10.1371/journal.pone.0130070

Moran, N., Jakubowski, K. & Keller, P. E. (2017). *Improvising duos - visual interaction collection, 2011* [moving image]. University of Edinburgh. https://doi.org/10.7488/ds/2153

Neuhoff, H., Polak, R. & Fischinger, T. (2017). Perception and evaluation of timing patterns in drum ensemble music from Mali. *Music Perception*, *34*(4), 438–51. https://doi.org/10.1525/mp.2017.34.4.438

Polak, R. (2010). Rhythmic feel as meter: Non-isochronous beat subdivision in jembe music from Mali. *Music Theory Online*, *16*(4). https://doi.org/10.30535/mto.16.4.4

Polak, R. (2017). The lower limit for meter in dance drumming from West Africa. *Empirical Musicology Review*, *12*(3-4), 205–26. https://doi.org/10.18061/emr.v12i3-4.4951

Polak, R. & London, J. (2014). Timing and meter in Mande drumming from Mali. *Music Theory Online*, *20*(1). https://doi.org/10.30535/mto.20.1.1

Polak, R., London, J., & Jacoby, N. (2016). Both isochronous and non-isochronous metrical subdivision afford precise and stable ensemble entrainment: A corpus study of Malian jembe drumming. *Frontiers in Neuroscience*, *10*, 285. https://doi.org/10.3389/fnins.2016.00285

Polak, R., Tarsitani, S., & Clayton, M. (2018). *IEMP Malian Jembe*. https://doi.org/10.17605/OSF.IO/M652X

Poole, A. I. (2013). *Groove in Cuban Dance Music: An Analysis of Son and Salsa*. [Unpublished doctoral thesis]. The Open University, U.K. https://oro.open.ac.uk/61186/1/680519.pdf

Poole, A., Tarsitani, S. & Clayton, M. (2019). *IEMP Cuban Son & Salsa*. https://doi.org/10.17605/OSF.IO/SFXA2

Rocamora, M., Cancela, P., & Biscainho, L. W. P. (2019). Information theory concepts applied to the analysis of rhythm in recorded music with recurrent rhythmic patterns. *Journal of the AES*, *67*(4), 160–173. https://doi.org/10.17743/jaes.2019.0003

Rocamora, M., Jure, L., Marenco, B., Fuentes, M., Lanzaro, F. & Gomez, A. (2015). An audio-visual database of Candombe performances for computational musicological studies. *II Congreso Internacional de Ciencia y Tecnología Musical (CICTeM),* Buenos Aires (pp. 17–24). http://dedicaciontotal.udelar.edu.uy/adjuntos/produccion/1658_academicas__academicaarchivo.pdf

Roeske, T.C., Tchernichovski, O., Poeppel, D. & Jacoby, N. (2020). Categorical rhythms are shared between songbirds and humans. *Current Biology*. https://doi.org/10.1016/j.cub.2020.06.072

Serra, X. (2014). Creating research corpora for the computational study of music: The case of the CompMusic Project. *AES 53rd International Conference: Semantic Audio*. 27-29 January 2014, London, UK. https://pdfs.semanticscholar.org/aff2/7e37a220d62727333b1455e54de1fb856cd7.pdf

Sloetjes, H. & Wittenburg, P. (2008). Annotation by category: ELAN and ISO DCR. *Proceedings of the 6th International Conference on Language Resources and Evaluation*. Marrakech, Morocco, 28–30 May 2008 (pp. 816–820). https://www.aclweb.org/anthology/L08-1034/

Tan, S. (2008). Returning to and from "innocence": Taiwan aboriginal recordings. *The Journal of American Folklore*, *121*(480), 222–235. https://doi.org/10.1353/jaf.0.0005

Wilkinson, M., Dumontier, M., Aalbersberg, I. et al. (2016). The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. https://doi.org/10.1038/sdata.2016.18

Zemp, H. (1996). The/an ethnomusicologist and the record business. *Yearbook for Traditional Music*, *28*, 36–56. https://doi.org/10.2307/767806

APPENDIX A

TECHNICAL VALIDATION

Metrical Annotation

Manual annotation of metrical structure is not only itself an important aspect of IEMPDC, but also informs the assignment of onset times to metrical positions and is therefore of crucial importance. In each case we relied on metrical annotations made by the curators. In order to check the replicability of these annotations, we compared independent annotations of metrical structure in one example from each corpus. The results of this comparison are reported in this section. Annotators, who were either curators or were briefed by experts in the musical style, were instructed to produce a mark-up of the metrical structure of the piece by providing times for each metrical boundary (i.e., beat 1 of each cycle or measure). This was accomplished by tapping to the music in Sonic Visualiser and followed by manual correction if necessary. Contextual information was provided, e.g., the name of the meter or a rhythmic pattern to listen out for to help identify the metrical boundaries.

Beat positions are not objectively verifiable points in time, since they represent individual listeners' perceptions of the metrical structure: as Danielsen (2018) suggests, it may be more useful to regard beat positions as “bins” of finite width. The spread of taps on each beat gives an indication of the accuracy with which these “bins” can be estimated by tapping: SDs range from about 9-28 ms (0.3-2.5% of the mean inter-tap interval, see Table A1).

Table A1. Variability in annotation of metrical structure.

Corpus	Sample	Annotators	Number of tapped beats	SD$_{\text{Difference}}$ (ms)[a]	SD$_{\text{Difference}}$ (ms) / Mean tap interval[b]
CSS	Song 2	AP, MC, LL	100	9.2	0.3%
ESQ	Haydn take 2	MC, LL, KJ	253	28.1	2.5%
MJ	Maraka 1	RP, MC, LL	123	20.5	1.4%
NIR	PrB_Jhinjhoti_Rupak	MC, LL, ST	166	27.9	0.6%
TS	3 Bousaadeya	RJ, MC, LL	451	12.5	1.5%
UC	Take 211	LJ/MR, MC, LL	85	18.8	0.8%

Note. CSS = Cuban Son & Salsa, ESQ = European String Quartet, MJ = Malian Jembe, NIR = North Indian Raga, TS = Tunisian Stambeli, UC = Uruguayan Candombe. One section from each corpus was tapped by three different annotators.

[a] Measures are SD of the mean absolute difference of tap times from the mean position.

[b] Measures are the ratio of this figure to the mean inter-tap interval.

Onset Extraction and Assignment

Event onset calculation used parameters set either by automatic validation or trial and error (see paper, ‘Onset extraction (Onsets_Raw)’), with the aim of achieving accurate estimation of most onsets with a low proportion of false positives. Errors were estimated by manually checking extracts of audio files (16 samples of approximately 1 minute each, including at least 2 for each corpus). The mean proportion of event onsets missed was 4% (range 0-23%), and the mean proportion of false positives was 6.5% (range 1-17%). The larger figures in each case relate to a sample of tabla (where quiet onsets were missed: the great dynamic range meant that capturing the quietest onsets would have produced unmanageable numbers of false positives), Indian slide guitar and cello (melodic instruments with a variety of ways of articulating notes and of timbres). The lowest error rates (<2%) are found for percussive instruments with relatively small dynamic range and recordings with little bleed from other instruments.

Based on a reliable estimation of the metrical structure, event onset times were extracted and assigned to metrical positions (up to 32 per cycle depending on structure and tempo), according to the procedure described above. Since this was carried out by the curators of MJ and UC using procedures independent of those used by Clayton across IEMPDC, we compare different versions of the selected onset data for two recordings, one each from these corpora: MJ_Maraka1 and UC_211.

- *Onset detection.* We detected event onsets for each sample track using the script shared in Technical resources/Onset detection, selecting parameters manually to ensure almost all drum onsets were captured but there were few false positives. Independently, for MJ_Maraka1, Soundforge Pro 10 (Sony) and Wavelab 7 (Steinberg) were used for automatic onset detection and marking of the Jembe 2 and Dundun parts. Onsets then were individually checked and corrected manually, where necessary, to the beginning of events immediately before the first peak in the amplitude. The lead drum (Jembe 1) was marked manually from scratch in Cubase (Steinberg). For UC_211, a standard onset detection algorithm was used based on the Spectral Flux (see corpus Documentation file for more detail). The resulting events were manually validated and/or corrected, by inspecting the audio and video files.

- *Onset assignment.* Onset assignment was carried out independently of the curators by calculating intermediate metrical positions as equal portions of each cycle [3] and selecting the closest onset, as long as it fell within a window of +/-60ms (UC_211) or +/-45ms (MJ_Maraka1, which accelerates to a faster tempo and therefore requires a smaller window). These onset assignments were manually checked by importing into SV and auditioning, to identify false positives (e.g., due to bleed between tracks), or onsets not intended to fall on the defined metrical positions (such as triplet figures where three strokes are played in the time of two beats or subdivisions).

In this way two independent sets of assigned onsets can be compared:

- CUR That produced by the corpus curators.
- MAN That produced by Clayton after manual checking.

The next two sub-sections address two key questions: (i) how close are the patterns of onset assignment (do they find onsets on the same metrical positions), and (ii) how close are the estimated onset times in the two sets of data?

ONSET ASSIGNMENTS

For each instrument in the two sample pieces, what percentage of onsets from each set are also present in the other? The comparison shows that the two sets are very closely aligned, with a mean of 3% of onsets selected in only one set across all of the instruments (see Table A2).

Two instruments stand out as having higher rates of mismatch: the Jembe 1 part in MJ_Maraka1 and the Repique part in UC_211. Both ensembles comprise one lower-pitched drum (Dundun/Piano), a higher-pitched drum which plays repetitive patterns (Jembe 2/Chico), and a higher-pitched drum that plays highly varied "lead" patterns (Jembe 1/Repique).

Table A2. Comparison of onset assignment using two independent methods, using sample pieces from the MJ and UC corpora.

	MJ_Maraka1			UC_211		
	Jembe 1	Jembe 2	Dundun	Repique	Chico	Piano
Onsets assigned only in MAN	39/995 (3.9%)	1/961 (0.1%)	4/701 (0.6%)	38/775 (4.9%)	17/1295 (1.3%)	25/549 (4.6%)
Onsets assigned only in CUR	84/1040 (8.1%)	0/960 (0%)	0/697 (0%)	91/828 (11.0%)	92/1280 (0.2%)	3/527 (0.6%)

It is the Jembe 1 and Repique that take most of the attention in manually checking onset assignments, since they can include features such as rolls, flams, and cross-rhythmic variations. The figures here show that in these two cases, Clayton (as a non-expert in the specific style and repertory) was somewhat more conservative, removing some onsets that the specialist curators felt more confident about assigning to metrical positions. Overall, this comparison shows that with relatively simple and unambiguous patterns, different methods can produce almost identical onset assignments. With more complex patterns, more points of disagreement emerge, but agreement is nonetheless >89%.

ONSET TIMINGS

For each of the instruments, means and confidence intervals were computed of the difference between onsets of the MAN and CUR sets on the same metrical positions. The small positive values for the means indicate that the IEMPDC Onset detection script tends to calculate slightly later onset times than those corrected visually. The biggest difference is for instruments with the lowest pitch in each ensemble (Dundun and Piano, mean difference 4.86 and 3.70 ms respectively, see Table A3). Again, the two independent datasets are closely matched.

Table A3. Mean difference (ms) in estimated drum onset times using two independent methods, using sample pieces from the MJ and UC corpora.

	MJ_Maraka1			UC_211		
	Jembe 1	Jembe 2	Dundun	Repique	Chico	Piano
Mean difference (ms) between MAN and CUR sets	0.57	1.24	4.86	2.67	3.63	3.70
95% CI (ms)	[0.43, 0.71]	[1.18, 1.30]	[4.56, 5.15]	[2.37, 3.10]	[3.38, 3.88]	[3.46, 3.93]

Movement Data

Movement data extracted using the EyesWeb Optical Flow patch is included in the collection for parts of the NIR, MJ and CSS corpora. This process has been validated separately through comparison of EyesWeb output and motion capture data for similar corpora for which both video and motion capture data were available. Correlations between Motion Capture and Optical Flow data ranged between 0.75 and 0.98 (Jakubowski et al., 2017).

Manual Annotation of Interaction between Musicians

Manual annotations of bouts of interaction between performers is included for parts of the CSS and NIR corpora. These annotations are similar to those produced for Moran et al.'s Improvising Duos corpus (Moran & Keller, 2015; Clayton et al., 2017; Moran et al., 2017), analyzed in Eerola et al. (2018). These annotations were carried out by three observers and validation of these data is addressed elsewhere: in brief, time series of the three raters were matched using dynamic time warping, and analysis of interrater agreement on the time-adjusted data resulted in an average κ (Cohen's Kappa) of 0.797 ($Z = 13.8$, $p < .001$). Annotations included here for CSS and NIR were made by a single observer and have not been used to date for published analysis; they are included here as supplementary data and for use by other researchers, for example to compare with independent annotations.

APPENDIX B

VISUALIZATION OF MEDIA AND ANNOTATIONS

Media files can be previewed directly within OSF's interface. Annotation data can also be visualized using freely-available programs such as Sonic Visualiser and ELAN. Examples of visualizations are shown in Figures B1 and B2 below. Figure B1 illustrates the visualization of a selection of data from the NIR corpus in ELAN 5.5. This was created using the following steps:

1. Create a new ELAN file and File/Import/CSV or Tab-delimited Text File to open NIR_PrB_Jhinjhoti_2Gats_Annotation_Sample.csv; save the result as PrB_Sample_Tiers.eaf. The five columns should be imported as Tier, Begin Time, End Time, <deselect> and Annotation respectively.
2. Edit the file NIR_PrB_jhinjhoti_2Gats_Metre_Rupak_Sample.csv to create an 'end time' column to the right of the Time for each annotation; the end time should be the start time of the following annotation. Then import this csv file as in step A and save as PrB_Sample_Metre.eaf; the first three columns are mapped as Annotation, Begin Time and End Time respectively.
3. Use File/Merge Transcriptions… to combine the two Elan files, saving as PrB_Sample_Merged.eaf.
4. Change the name of Cycle to METRE (by right-clicking on the name and selecting Change Attributes of Cycle).
5. Use Edit/Linked Files to Add all of the files in the Sample/Media folder; NIR_PrB_Jhinjhoti_2Gats_Tabla_sample.wav is selected here in the dropdown box to the left of the audio waveform window. The Control window (shown top right) can be used to select which audio track(s) play back.
6. Adjust window sizes and resolution of timeline (font size of annotations adjusted here to 18pt for readability).

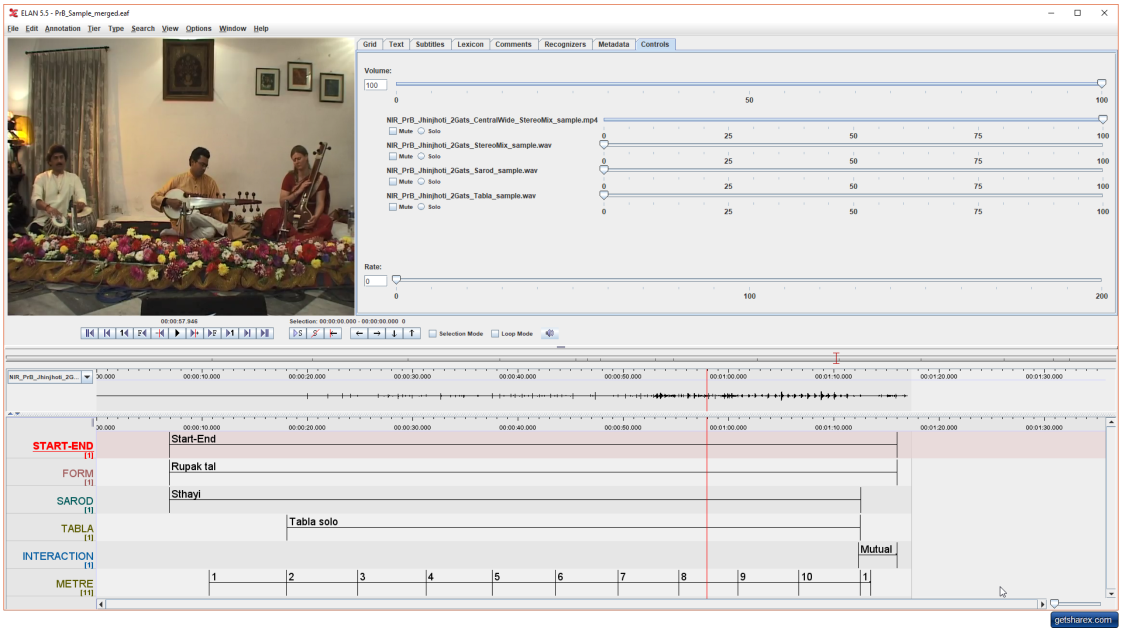
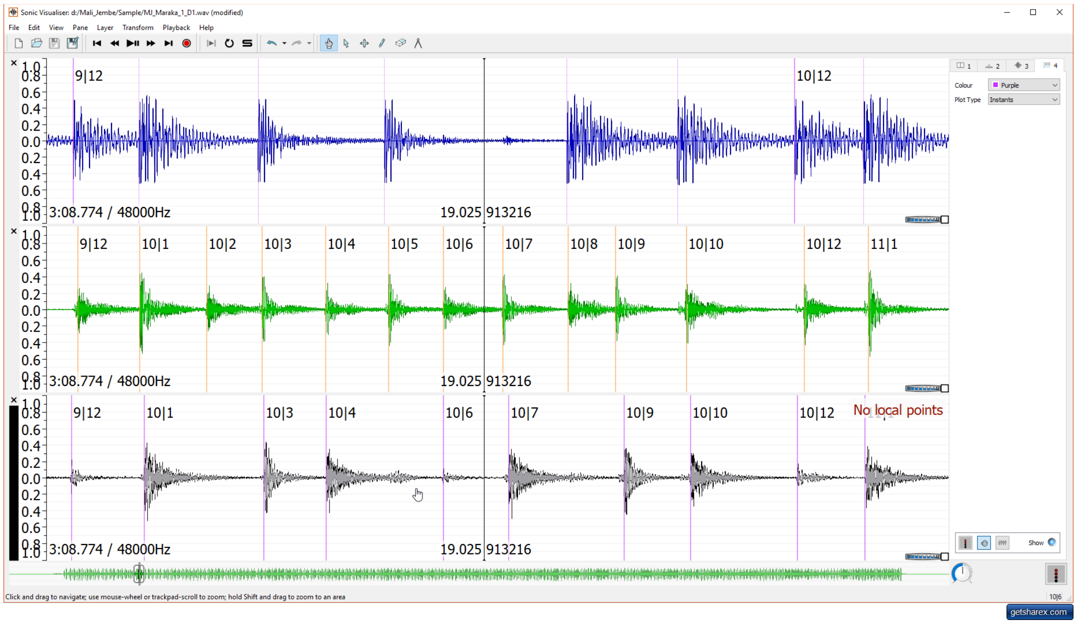

ELAN 5.5 - PrB_Sample_merged.eaf
File Edit Annotation Tier Type Search View Options Window Help
Grid Text Subtitles Lexicon Comments Recognizers Metadata Controls
Volume:
100
0 50 100
NIR_PrB_Jhinjhoti_2Gats_CentralWide_StereoMix_sample.mp4
□ Mute ○ Solo
NIR_PrB_Jhinjhoti_2Gats_StereoMix_sample.wav
□ Mute ○ Solo
NIR_PrB_Jhinjhoti_2Gats_Sarod_sample.wav
□ Mute ○ Solo
NIR_PrB_Jhinjhoti_2Gats_Tabla_sample.wav
□ Mute ○ Solo
0 25 50 75 100
Rate:
0
0 100 200
00:00:57.946
Selection: 00:00:00.000 - 00:00:00.000 0
□ Selection Mode □ Loop Mode
NIR_PrB_Jhinjhoti_2G...
00.000 00:00:10.000 00:00:20.000 00:00:30.000 00:00:40.000 00:00:50.000 00:01:00.000 00:01:10.000 00:01:20.000 00:01:30.000
START-END [1]
Start-End
FORM [1]
Rupak tal
SAROD [1]
Sthayi
TABLA [1]
Tabla solo
INTERACTION [1]
Mutual
METRE [11]
1 2 3 4 5 6 7 8 9 10 1
getsharex.com

Fig. B1. Visualization of sample data from the NIR corpus in ELAN 5.5.

Figure B2 illustrates the visualization of a selection of data from the MJ corpus in Sonic Visualiser 3.2.1 (SV). This was created using the following steps:

1. Open MJ_Maraka_1_Onsets and save a local copy with a new name; delete all columns except Label SD, J1, J2 and D1.
2. Open audio file MJ_Maraka_1_D1.wav in Sonic Visualiser.
3. Use File/Import Annotation Layer to import the onset labels for this file. Open the CSV file created in A, select column Label SD as 'Label', column D1 as 'Time' and all other columns as <ignore>.
4. Using File/Import More Audio, import the other two tracks and their corresponding annotations.
5. Use mouse wheel to adjust time resolution until onset labels are visualized. The waveforms are now displayed with annotated labels at the detected onset times. (Nb. SV can create errors or omissions in the display of labels, as in D1 here, especially when using large files. The font size has been increased to 18 pt for readability.)
6. Save Session.



Colour: Purple
Plot Type: Instants

Click and drag to navigate; use mouse-wheel or trackpad-scroll to zoom; hold Shift and drag to zoom to an area

Fig. B2. Visualization of sample data from the MJ corpus in Sonic Visualiser 3.2.1 (SV).