

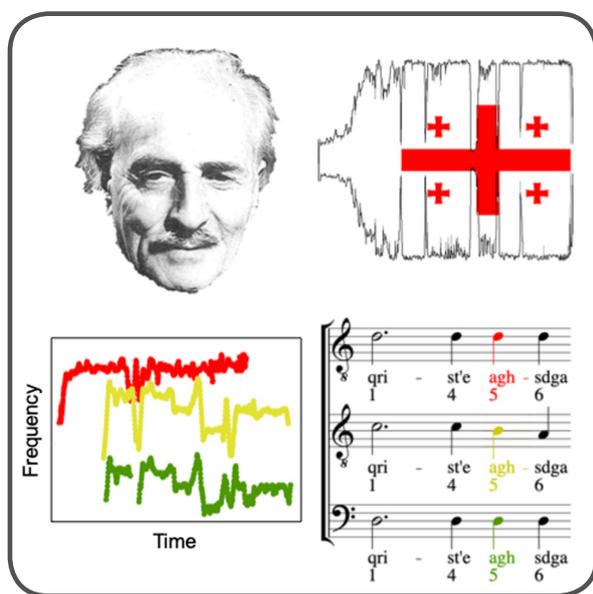
## DATASET

# Erkomaishvili Dataset: A Curated Corpus of Traditional Georgian Vocal Music for Computational Musicology

Sebastian Rosenzweig\*, Frank Scherbaum†, David Shugliashvili‡, Vlora Arifi-Müller\* and Meinard Müller\*

The analysis of recorded audio material using computational methods has received increased attention in ethnomusicological research. We present a curated dataset of traditional Georgian vocal music for computational musicology. The corpus is based on historic tape recordings of three-voice Georgian songs performed by the former master chanter Artem Erkomaishvili. In this article, we give a detailed overview of the audio material, transcriptions, and annotations contained in the dataset. Beyond its importance for ethnomusicological research, this carefully organized and annotated corpus constitutes a challenging scenario for music information retrieval tasks such as fundamental frequency estimation, onset detection, and score-to-audio alignment. The corpus is publicly available and accessible through score-following web-players.

**Keywords:** Georgian Vocal Music; Artem Erkomaishvili; Score-Audio Annotations



**Figure 1:** Erkomaishvili dataset with annotations.<sup>1</sup>

## 1. Introduction

Georgia has a rich cultural heritage. Its traditional polyphonic vocal music, which has been acknowledged as Intangible Cultural Heritage by the UNESCO in 2001,

is one of the most prominent examples. Being an orally transmitted singing tradition, most of the sources are available as field recordings. Ethnomusicological research is usually conducted on the basis of notated musical scores, which are obtained by manually transcribing the audio material. Such approaches are problematic since important tonal cues and performance aspects are likely to get lost in the transcription process. Consequently, the analysis of recorded audio material using computational methods has become increasingly important in musicological research (Ganguli and Rao, 2018; Serra, 2014a; Müller et al., 2009).

With the goal to contribute to the preservation of the Georgian cultural heritage and to support research on Georgian vocal music, we have created a manually annotated dataset of traditional three-voice Georgian songs. The corpus is based on recordings of the former Georgian master chanter Artem Erkomaishvili, which were recorded in 1966 with tape recorders by the ethnomusicologist Kakhi Rosebashvili. The original recordings are preserved in the archive of the Georgian Folk Music Department of the Tbilisi State Conservatoire. 101 recordings are publicly available.<sup>2</sup> Due to a lack of fellow singers, Artem Erkomaishvili sung all three voices on his own, which was made possible through a three-stage overdubbing recording process. Beyond the historic recordings, there exist transcriptions of all songs in Western staff notation created by the Georgian ethnomusicologist David Shugliashvili (Shugliashvili, 2014). Furthermore, Müller et al. (2017) annotated the three-part recording structure and fundamental frequency (F0) trajectories for the three voices in all recordings.

Our main contributions to the Erkomaishvili dataset are threefold. First, we have collated existing audio data

\* International Audio Laboratories Erlangen, DE

† University of Potsdam, Potsdam, DE

‡ Tbilisi State Conservatoire, GE

Corresponding author: Sebastian Rosenzweig  
(sebastian.rosenzweig@audiolabs-erlangen.de)

and annotations and introduced a uniform filename convention. Second, based on the existing transcriptions, the sheet music was brought into digital, machine-readable MusicXML-format. Subsequently, we manually annotated note onsets of the first voice in each of the recordings. This step has been carried out by an experienced annotator with the advice of domain experts. Third, in order to provide a direct and convenient access to the dataset, we developed an interactive web-based interface with score-following audio players that make use of the annotated data. Complementing the publicly available audio material, we release all fundamental frequency annotations, recording structure annotations, and note onset annotations along with this article. Additionally, we make the MusicXML-files of the symbolic transcripts publicly available.

Due to the importance of Artem Erkomaishvili's recordings for ethnomusicological research, the presented corpus is a vital source for studying tonal organization, intonation and harmonic and melodic thinking in traditional Georgian vocal music. Furthermore, the dataset can be used for developing and testing algorithms for music information retrieval (MIR) tasks such as fundamental frequency estimation, onset detection, or score-to-audio alignment.

The remainder of this article is organized as follows. First, we highlight related corpora and open-source tools for computational ethnomusicology (Section 2). Then, we give an introduction to traditional Georgian vocal music and explain the importance of Artem Erkomaishvili's recordings (Section 3). Subsequently, we provide detailed descriptions of the Erkomaishvili recordings, available transcriptions, and annotations (Section 4). Furthermore, we give an overview of the interactive web-based interface for accessing the dataset (Section 5). Finally, we sketch possible applications of the dataset for musicology and MIR research (Section 6).

## 2. Related Datasets and Tools

The goal of increasing reproducibility and transparency of scientific results has led to the release of various open datasets and open-source software for computational musicology. In the following, we want to give a short summary on related datasets and tools that are fundamental to our work on the Erkomaishvili dataset. One of the most extensive databases for computational ethnomusicology has been collected within the CompMusic research project (Serra, 2014b). The collection comprises recordings of Indian Art music (Carnatic and Hindustani music) (Srinivasamurthy et al., 2014), Turkish-Makam (Uyar et al., 2014; Dzhambazov et al., 2016; Şentürk, 2016), Jingju (Repetto and Serra, 2014; Gong et al., 2017), and Andalusian music (Repetto et al., 2018). The individual corpora, which include annotations of lyrics, scores, and editorial metadata, are hosted on the web-platform Dunya.<sup>3</sup> Kroher et al. (2016) released the corpus COFLA, a dataset for the computational study of Flamenco music. The Meertens Tune Collections<sup>4</sup> is a corpus of Dutch folk song recordings accompanied with syllabified lyrics, key annotations, phrase annotations, and transcriptions (van Kranenburg et al., 2019). Furthermore, the *Polyphony Project*<sup>5</sup> hosts a collection of

Ukrainian folk music recordings which is accessible via a web-based interface with multitrack audio and video players. A detailed overview of corpora for computational ethnomusicology can be found in (Panteli, 2018).

As the number of public datasets increases, so does the number of open-source toolboxes for computational analysis of music recordings. Prominent examples are LibROSA (McFee et al., 2015), Essentia (Bogdanov et al., 2013), MIR-Toolbox (Lartillot and Toiviainen, 2007) and Marsyas (Tzanetakis, 2009). Furthermore, tools such as Praat (Boersma, 2001), Sonic Visualiser (Cannam et al., 2010) and Tarsos (Six et al., 2013) offer graphical user interfaces to compute and display analysis results. Recently, a collection of implementations, mathematical descriptions and explanations of music processing algorithms with emphasis on didactic aspects was released (Müller and Zalkow, 2019).

## 3. Traditional Georgian Vocal Music

Despite its small size, Georgia is home to diverse singing traditions, which form an essential part of its cultural identity. The disparity of polyphonic Georgian vocal music in comparison to Western music is—among other aspects—based on the abundant use of “dissonances” and on the fact that the music is not tuned to the 12-tone equal-tempered scale. While musicologists agree on the not equal-tempered nature of traditional Georgian vocal music, the particular nature of the traditional Georgian tuning is an ongoing topic of intense and controversial discussions (Erkvanidze, 2016; Tsereteli and Veshapidze, 2014; Scherbaum, 2016). A related aspect, which by some musicologists has been considered characteristic for Georgian singing, is the importance of harmonic intervals, which often goes along with a relaxed precision of melodic intervals, e.g., (Chokhanelidze, 2010; Scherbaum et al., 2017).

One key towards understanding these phenomena is the analysis of high-quality audio recordings. A recently released research corpus of traditional Georgian vocal music (Scherbaum et al., 2019) meets all the quality criteria for computational analysis and allows for a systematic investigation of more than 200 performances. However, with few exceptions, it only captures the current performance practice in Svaneti, a historic province in Georgia. Regarding historical field recordings, the known publicly available audio material is rather limited. This is true despite the fact that there have been considerable efforts to record traditional Georgian vocal music, starting with phonograph recordings more than 100 years ago. Unfortunately, many recordings from the early days of the last century have not survived the course of time. The audio files that have survived are mostly of insufficient quality for computational analysis. A notable exception are the 1966 tape recordings of Artem Erkomaishvili (1887–1967)—one of the last Georgian master chanters—which are considered today as “original masterpieces of Georgian musical thinking” (Shugliashvili, 2014, p. XXVII). A part of the recordings was manually remastered and published on CD (Jgharkava, 2016). Today, the recordings of Artem Erkomaishvili are very likely the oldest collection of Georgian chants of sufficient size and quality for computational studies.

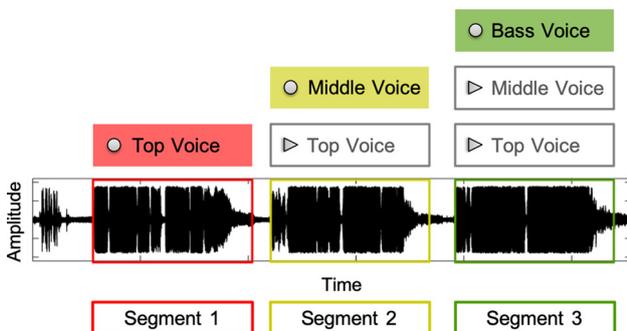
## 4. Erkomaishvili Dataset

In this section, we describe the main components of the Erkomaishvili dataset. More specifically, we first explain the specific recording procedure and elaborate on existing transcriptions (Section 4.1). Then, we give descriptions of the available manual annotations and the annotation process (Section 4.2). Finally, we present a semi-automatic method for the transfer of note onset annotations using alignment and interpolation techniques (Section 4.3).

### 4.1 Recordings and Transcriptions

In 1966, shortly before his death, Artem Erkomaishvili was asked to perform three-voice chants on his own by successively singing each of the individual voices. At the beginning of each recording, Artem Erkomaishvili announced the name of the song he was about to perform. After recording the top voice, one tape recorder was used to play back this first voice while a second tape recorder synchronously recorded the middle voice. Similarly, playing back the first and second voice, the bass voice was recorded, see **Figure 2**. In this way, Erkomaishvili accompanied his own recordings. However, due to this specific recording procedure, Artem Erkomaishvili usually began the middle and bass voices with a slight offset against the top voice. The resulting collection comprises 101 audio recordings with a total length of more than seven hours (see **Table 1**). Due to the distortions introduced by the tape recorders, the sound quality decreases with each recording stage. The strongest distortions typically occur in the third part, where it can sometimes be challenging to distinguish the bass voice from the other two voices. Additionally, since Artem Erkomaishvili was a bass singer, all songs are performed quite low. Considering the distortions and low-frequency content in the audio material, the recordings constitute a particularly challenging scenario for audio processing algorithms.

Transcriptions of Artem Erkomaishvili’s recordings in Western staff notation have been published in the book “Georgian Church Hymns, Shemokmedi School” by David Shugliashvili (Shugliashvili, 2014). The book



**Figure 2:** Illustration of three-stage recording process.

**Table 1:** Overview on Erkomaishvili’s recordings.

# Songs	Total Duration (hh:mm:ss)	Mean/Min/Max Duration (mm:ss)
101	7:04:49	04:12/00:40/13:37

contains 118 consecutively numbered transcriptions with song titles given in Georgian and English language, and song lyrics in Georgian and Latin letters. As opposed to Artem Erkomaishvili’s performances, the transcriptions are notated in a higher register to account for a wider singer audience. During curation of the dataset, we used the score numbers in the book as unique file identifiers (Georgian Chant Hymns-IDs, abbr. GCH-IDs). As a naming convention, we included the GCH-IDs as three digit prefix consistently in all audio, sheet music, and annotation filenames. Since the publicly available audio collection comprises only 101 recordings, the Erkomaishvili dataset does not contain data for the following GCH-IDs: 021, 028, 037, 038, 039, 055, 064, 075, 082, 084, 096, 117, 118. Furthermore, recordings with the following GCH-IDs include two songs (second song in brackets): 022 (023), 043 (044), 058 (059), 102 (103).<sup>6</sup>

### 4.2 Manual Annotations

In this section, we explain all manual annotations contained in the Erkomaishvili dataset. From a previous study, we included recording structure annotations (Section 4.2.1) and semi-automatically annotated F0-trajectories of the three voices (Section 4.2.2). As one main contribution of this article, we generated digital sheet music (Section 4.2.3) and onset annotations (Section 4.2.4) with the help of an experienced annotator. In the following, we use the song “Da Sulisatsa” (GCH-ID 087) as a running example.

#### 4.2.1 Segment Annotations

As explained in Section 4.1, the first voice appears three times in every recording and marks the beginning and end of each recording stage. Due to varying tape velocities, the durations of second and third stages may slightly deviate from the duration of the first stage. However, for most of the recordings with few exceptions (GCH-ID 004, 015, 107), it is a good approximation to assume the same duration for all three stages. Following this assumption, Müller et al. (2017) determined in all recordings the positions of three segments with equal duration (see **Figure 2**). Thereby, the segment start is defined by the start of each recording stage, whereas the segment duration is defined by the duration of the first stage. The segment annotations are available in CSV-format and contain six time-stamps corresponding to the start and end positions of the three segments.

#### 4.2.2 Fundamental Frequency Annotations

As part of the same study on Artem Erkomaishvili’s recordings, Müller et al. (2017) annotated F0-trajectories of the three voices for all 101 songs using a semi-automatic tool with a graphical user interface. The annotation procedure was as follows: first, the user specified temporal-spectral constraint regions in an enhanced time–frequency representation of the recording. Subsequently, F0-trajectories were automatically computed within the specified regions using an F0-estimation algorithm similar to Melodia (Salamon and Gómez, 2012). In this way, the annotator could guide the estimation process. Additionally, the tool provides audiovisual feedback mechanisms for

validation purposes and allows for correcting the computed F0-trajectories. The resulting annotated trajectories have a time resolution of 5.8 ms and a log-frequency resolution of 10 cents. The two-column annotation files in CSV-format contain equally-spaced timestamps in seconds in the first column and the F0-estimates in Hertz in the second column. The value of 0 Hz is used to indicate parts where the voice is inactive. Since the F0-trajectories of the three voices were annotated independently from the segment annotations, a few F0-values might be annotated outside the segment boundaries. The F0-trajectories for our running example, plotted within the segment boundaries on a logarithmic frequency axis, are depicted in **Figure 3a**. The activations of the F0-trajectories are shown in **Figure 3b**.

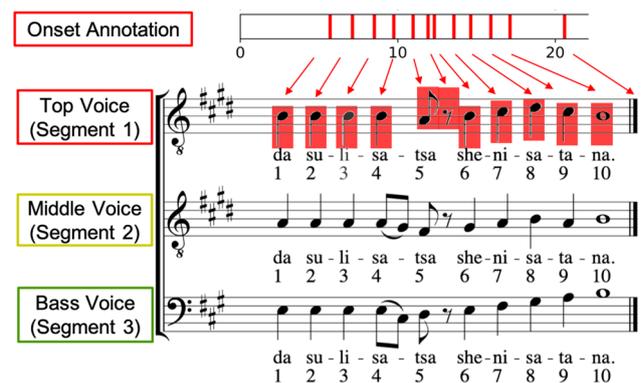
#### 4.2.3 Digital Sheet Music

Computational comparisons of the transcribed musical scores with the actual performances of Artem Erkomaishvili require digital scores in machine-readable format. One way to transfer printed sheet music to digital formats is to use Optical Music Recognition (OMR) systems (Byrd and Simonsen, 2015; Rebelo et al., 2012). However, despite the advances over the last years, such systems are still error-prone and usually require labor-intensive manual corrections to obtain good quality results. Furthermore, most systems are not able to recognize characters from the Georgian writing system that are contained in the lyrics of the scores. Due to these circumstances, the transcriptions of David Shugliashvili were manually transferred to digital scores in MusicXML-format using the scorewriter programs Finale<sup>7</sup> and Sibelius.<sup>8</sup> As opposed to Western music, the traditional Georgian songs do not have a fixed musical time signature and are not organized using measures. However, a musical reference grid is beneficial for orientation within the scores. Furthermore, it helps to align the audio with the sheet music domain, as we will see in the following sections. Therefore, we introduce the concept of Quarter Note References (QNRs)—a concept of rather technical nature which has no further musical importance in traditional Georgian vocal music. A QNR is assigned to each note and indicates its position in terms of

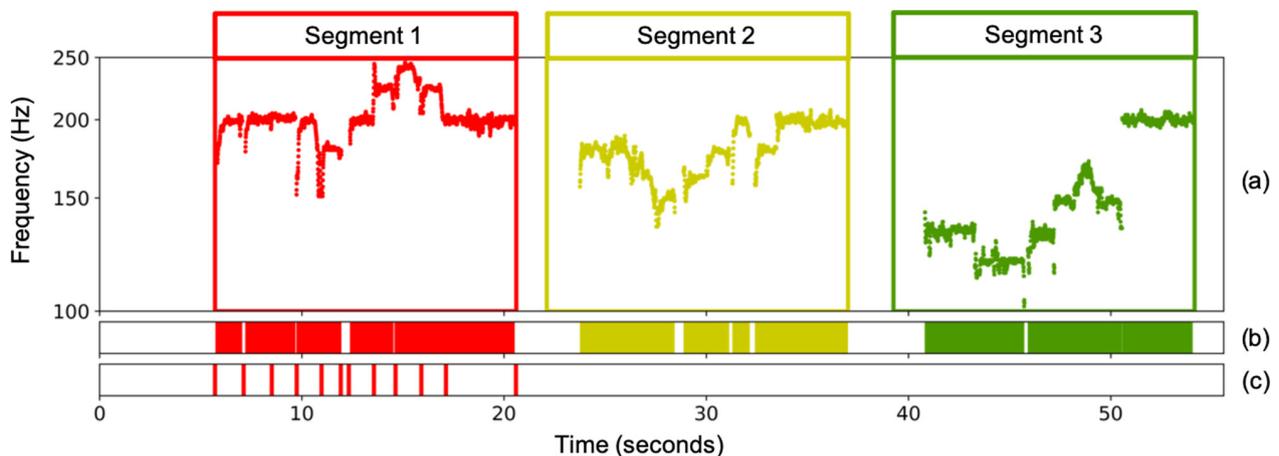
quarter notes from the beginning of the score. Following this concept, QNR 1 refers to the first note in the score, whereas QNR 2 refers to the note on the second quarter note (second beat), which is not necessarily the second note in the score. In this way, QNRs are assigned to notes in every system of the score. In case the system contains shorter notes than quarter notes (e.g., eighth notes), QNRs can be floating point numbers indicating fractions of quarter notes. For visualization purposes, only integer QNRs have been added to the lyrics of the individual voices using the music21 Python toolkit (Cuthbert and Ariza, 2010). The generated digital score with QNRs for our running example is depicted in **Figure 4**.

#### 4.2.4 Onset Annotations

In order to align the audio recordings with the digital scores, we manually annotated note and rest onset positions in the recordings using the open source software Sonic Visualiser (Cannam et al., 2010). Due to practical reasons, we only annotated note and rest onsets of the first voice. The onsets of the second and third voice were then derived from the onsets of the first voice and the segment annotations (for more details see Section 4.3).



**Figure 4:** Digital score for “Da Sulisatsa” (GCH-ID 087). The annotated note and rest onsets for the top voice are highlighted in red. The QNRs are displayed underneath the lyrics of each voice.



**Figure 3:** Illustration of available annotations for the song “Da Sulisatsa” (GCH-ID 087). **(a)** F0-trajectories within annotated segment boundaries plotted on a logarithmic frequency axis. **(b)** Activations of F0-trajectories. **(c)** Onset annotations including segment end.

**Figure 3c** depicts the onset annotations of the first voice, which complement the existing manual segment and F0-annotations. **Figure 4** shows the correspondences of the onset annotations to note events in the digital score for our running example. As a convention, the onset annotations include the end of the first voice (end of first segment) as a last timestamp.

### 4.3 Onset Computation

Generating onset annotations for the middle and bass voices in the Erkomaishvili recordings is challenging due to the polyphony and the poor audio quality in the second and third recording stages. In addition, Artem Erkomaishvili's low voice and the Georgian singing style with the abundant use of pitch slides in the beginning, end, and in between consecutively sung notes complicates this task. Therefore, instead of manually annotating the onsets of the middle and bass voices, we computed the onsets using a semi-automatic approach. As described earlier, due to the overdubbing recording process, the top voice is played back in the second and third segment and serves as reference for the other two voices. Using the segment annotations from Section 4.2.1, we mapped the onset annotations of the top voice to the other two segments by calculating the difference between the segment start positions and adding it to the top voice onset timestamps. In a subsequent step, we determined the onsets of the middle and bass voices using the previously introduced QNR grid. For notes of the middle and bass voices that share the same QNR as notes in the top voice (two notes that are exactly on the same score time), we assigned the mapped onset time of the top voice note. In order to obtain onsets of notes with a unique QNR (such as the notes between QNR 4 and QNR 5 in the middle and bass voices of our running example in **Figure 4**), we interpolated between the neighboring note onsets according to the QNR grid. We want to note that this approach requires the segments to be of equal duration and the tape velocity to stay constant during all recording stages in order to obtain a close approximation of the onsets for the second and third voices. Furthermore, the three voices are required to be sung in sync. These requirements can be assumed for most of the songs in the dataset. However, outliers can be found in the recordings with identifiers GCH-ID 004, 015, and 107. These recordings suffer from strong tape recorder artifacts. Therefore, it would be necessary to manually correct playback velocity and pitch prior to onset computation. In our Erkomaishvili dataset, we want to preserve the original recordings while indicating cases where outliers occur. We leave further modifications of the historic audio material to future studies. For all 101 recordings, the annotated onsets for the first segments and the computed onsets for the second and third segments are released in CSV-format (one CSV-file per segment). In the CSV-files, each row contains information for one onset in the following format: onset index, onset time in seconds relative to the segment start, onset "end" in seconds relative to the segment start (equivalent to the onset time of the next onset), QNR of the corresponding note or rest, QNR of the next note or rest.

## 5. Web-Based Interface

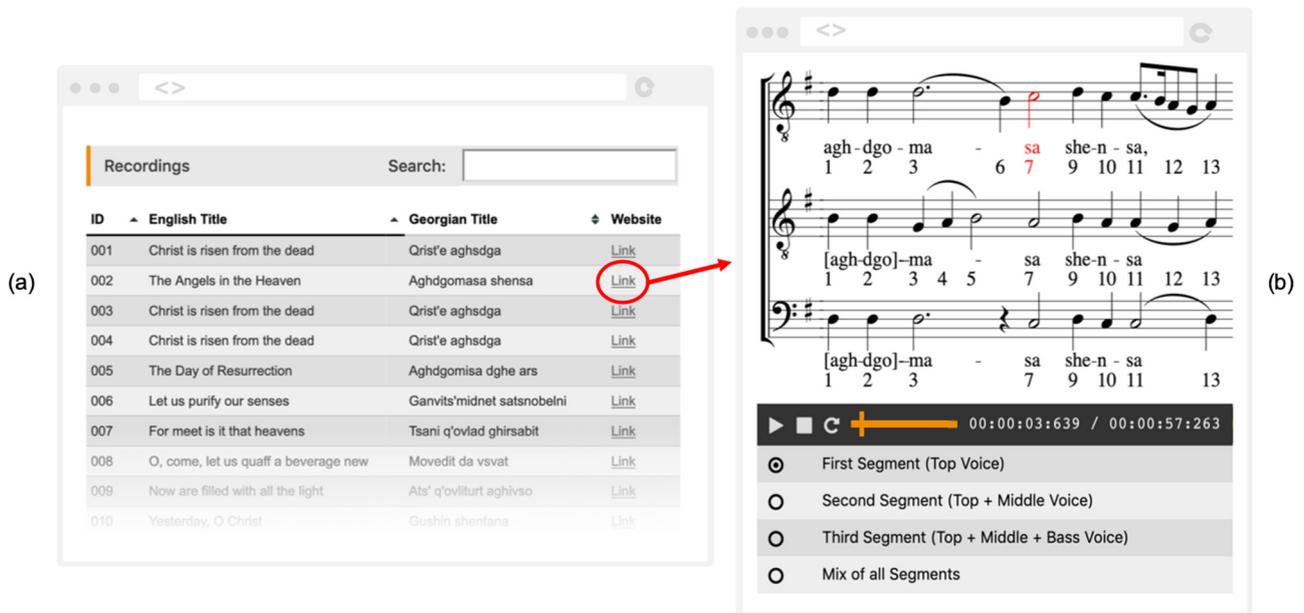
The public availability of MIR research corpora is essential for the reproducibility of scientific results, as well as for the preservation and dissemination of audio material and its annotations. Platforms such as Zenodo<sup>9</sup> offer to publicly share and distribute scientific data, while also providing citeable Digital Object Identifiers (DOIs). However, the interdisciplinary field of computational musicology requires platforms beyond data repositories, which support a cross-disciplinary scientific exchange by offering a direct, intuitive, and comprehensive access to the data. This can be accomplished by means of interactive interfaces that bridge the gap between the musicological and the audio domain (e.g., see Gasser et al., 2015; Jeong et al., 2017; Röwenstrunk et al., 2015).

As one main contribution of this article, we developed a publicly accessible web-based interface<sup>10</sup> which hosts the full dataset. The interface provides download links to all segment, fundamental frequency, and onset annotations. Each song in the dataset has its individual sub-page, which is accessible through an interactive table with search and sorting functionalities as shown in **Figure 5a**. The central element of each sub-page is a multitrack audio player (Werner et al., 2017) with score-following functionality (Zalkow et al., 2018). The displayed digital sheet music (given as an MEI file) is dynamically rendered in the web-browser with the help of Verovio (Pugin et al., 2014). The user can seamlessly switch between the three individual recording segments and a mix version of the three segments. In parallel, sung notes, lyrics, and QNRs are highlighted in the score according to the manually annotated onsets of the top voice and the automatically generated onset annotations for the middle and bass voices (see **Figure 5b**). In summary, beyond providing a non-technical and multimodal access to the Erkomaishvili dataset, the developed interface constitutes a first application scenario based on our annotations.

## 6. Applications for MIR and Musicology

The Erkomaishvili dataset can be used to address a wide range of research questions including technical as well as musicological ones. For example, a cappella vocal music is a challenging scenario for various MIR tasks such as F0-estimation (Salamon et al., 2014), onset detection (Böck et al., 2012), and score-to-audio alignment (Thomas et al., 2012; Arzt, 2016; Müller et al., 2019). In particular, the not equal-tempered nature of the Georgian songs and the characteristic pitch slides in traditional Georgian singing constitute challenging test scenarios for MIR algorithms. The Erkomaishvili dataset is one of few publicly available datasets on polyphonic a cappella singing (Cuesta et al., 2018; Scherbaum et al., 2019). Due to the overdubbing procedure, the audio material provides a suitable scenario for studying source separation (Cano et al., 2019), audio segmentation (Rosenzweig, 2017), and audio restoration techniques (Godsill et al., 2002).

Computational ethnomusicology is a rather young and still evolving field of research (Tzanetakis et al., 2007; Gómez et al., 2013; Tzanetakis, 2014). Its potential depends strongly on the existence of data collections



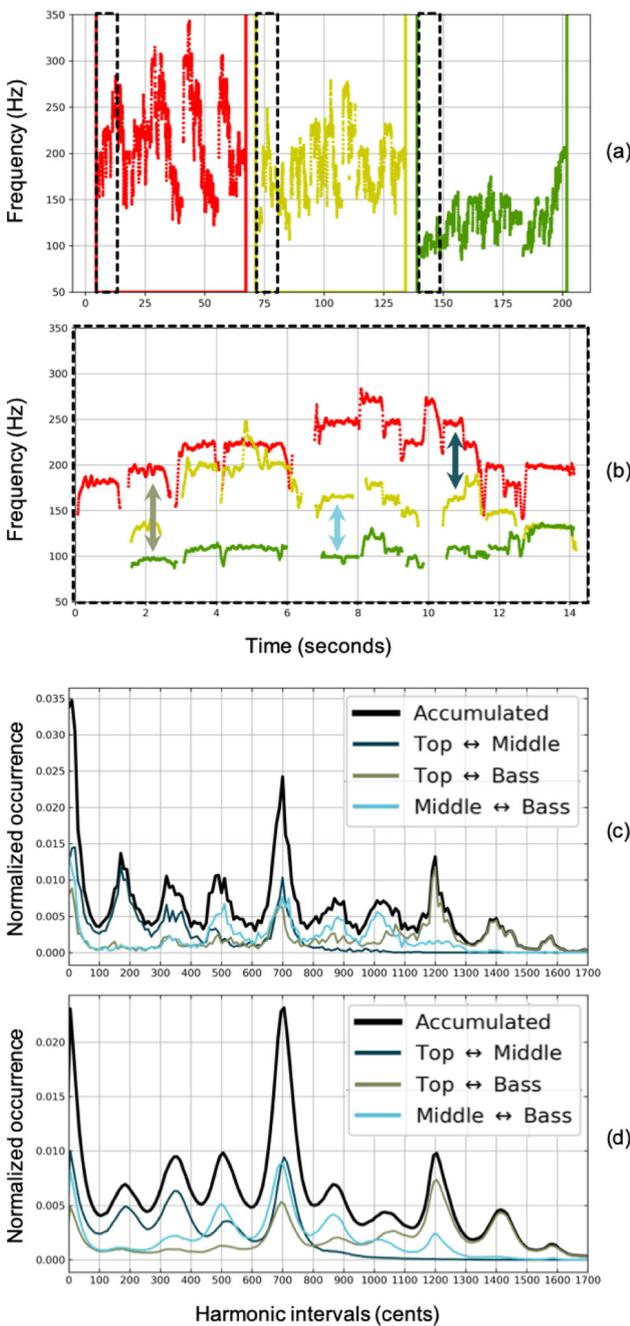
**Figure 5:** Web-based interface for accessing the Erkomaishvili dataset. (a) Main page with overview table. (b) Sub-page for the song “Aghdgomasa shensa” (GCH-ID 002) with score-following player.

which on the one hand are musically relevant, and on the other hand are of sufficient quality for the application of computational tools. The presented corpus meets both of these criteria. Its musicological relevance is undisputed. Ethnomusicologist John Graham, for example, writes: “Any theory must account for both the tuning system heard in the 1966 Erkomaishvili recordings and evidence from earlier singers and other regional chant systems seen in the transcription record” (Graham, 2015, p. 292). Some musicologists even believe that only through the analysis of historical recordings (such as the Erkomaishvili collection), the Georgian musical system can be understood (Erkvanidze, 2016).

In the following, we illustrate the potential of our annotations in two case studies using the song “Gushin Shentana” (GCH-ID 010) as a running example. In the first case study (see **Figure 6**), we analyze the harmonic content of the Erkomaishvili recordings by computing distributions of sung harmonic intervals following the approach of Müller et al. (2017). To this end, the annotated F0-trajectories of the top, middle, and bass voices (see **Figure 6a**) are superimposed using the segment annotation (see **Figure 6b**). Then, for each time position, the intervals (given in cents) between the F0-trajectories of the top and middle voices, the top and bass voices, as well as the middle and bass voices are computed. Finally, integrating the occurrences of the different intervals over time, we obtain for each of the three cases an interval distribution (see **Figure 6c**). By computing and averaging such distributions over all 101 Erkomaishvili recordings, we obtain the distributions shown in **Figure 6d**. Besides the peak around 0 cents (unison), the accumulated distribution exhibits a prominent peak around 700 cents (fifth), which reflects the importance of the fifth interval in traditional Georgian vocal music. The peak at around 350 cents, located between the minor third (300 cents) and major third (400 cents), indicates the not equal-tempered

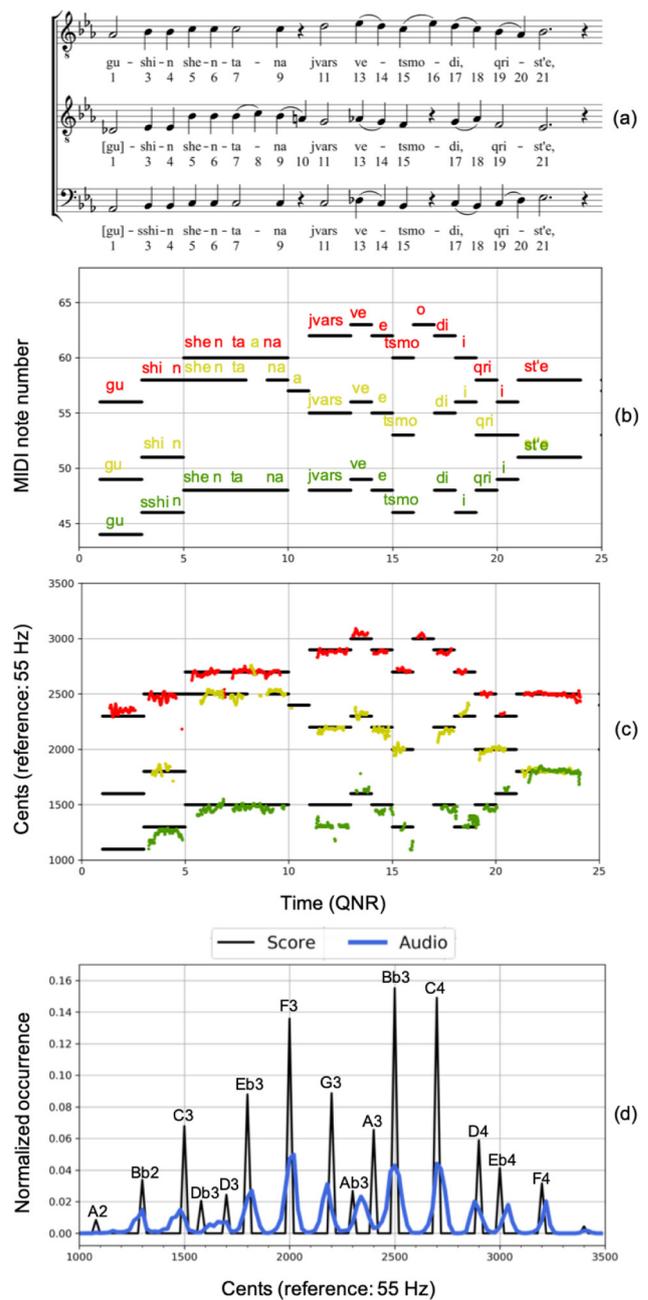
nature of traditional Georgian vocal music. For a more detailed study on traditional Georgian tuning, we refer to Scherbaum (2016).

What adds to the scientific value of the Erkomaishvili dataset is the availability of digital sheet music in Western staff notation for all songs (see Section 4.2.3). Although Western staff notation does not account for the not equal-tempered nature of traditional Georgian vocal music (see Section 3), the transcriptions can serve as a reference for more detailed studies on traditional Georgian tuning, e.g., as approximate guidance for MIR algorithms. Furthermore, qualitative comparisons with acoustical properties of Erkomaishvili’s recorded performances give insights into the challenges of transcribing not equal-tempered music. This is illustrated in our second case study (see **Figure 7**). In this study, we compare the pitch inventory as specified by the score representation with the pitch inventory as used by Artem Erkomaishvili. To this end, we proceed as follows. First, based on the digital sheet music (see **Figure 7a**), we generate a piano roll representation as shown in **Figure 7b**. Second, we extract stable regions in the F0-trajectories that roughly correspond to note events. For this task, we use an approach with morphological filters introduced by Rosenzweig et al. (2019). Third, we temporally align the filtered F0-trajectories with the piano roll representation using the onset annotations. By making use of the previously introduced QNR concept (see Section 4.2.3), we obtain a QNR axis for both the score and the audio information. As a common frequency axis, we choose a logarithmic axis in cents (reference frequency 55 Hz). Fourth, we adapt the filtered and aligned F0-trajectories to the piano roll representation using a global pitch shift. This step is necessary since the transcriptions are notated in a higher pitch range than Artem Erkomaishvili’s original performance (see Section 4.1). We determine the global pitch shift by computing the difference between



**Figure 6:** Computation of harmonic intervals. (a) F0-trajectories of “Gushin Shentana” (GCH-ID 010). (b) F0-trajectories of all three voices superimposed using segment annotation (zoom region). (c) Histogram of harmonic intervals for “Gushin Shentana” (GCH-ID 010). (d) Histogram of harmonic intervals averaged over all 101 songs of the dataset.

the mean pitch of the piano roll representation and the mean pitch of the adapted trajectories (considering the trajectories of all voices jointly). In this way, we determine for our running example a pitch shift of 282 cents. The piano roll representation superimposed with the filtered and shifted trajectories is depicted in **Figure 7c**. In most of the cases, the extracted stable trajectory regions match the note events in the piano roll representation. However, a few regions in the F0-trajectories were not detected as “stable” (e.g., for the middle voice at QNR 10). In other



**Figure 7:** Comparison of transcribed score representation and annotated F0-trajectories for “Gushin Shentana” (GCH-ID 010). (a) Sheet music representation (excerpt). (b) Piano roll representation of score with lyrics (excerpt). (c) Adapted F0-trajectories for all three voices restricted to stable regions. (d) Pitch histograms for piano roll representation and adapted F0-annotation. The note names are given (A4 = 440 Hz).

cases (e.g., for the bass voice between QNR 11 and 13), the F0-values differ from the piano roll representation. To get an overall view on these deviations, we integrate the occurrences of pitch values of the piano roll representation and the adapted F0-trajectories over time. The two resulting distributions (“audio” and “score”) are depicted in **Figure 7d**. In general, both distributions exhibit similar peak locations. However, there exist two peaks in the audio distribution that deviate substantially from the score distribution. The most significant deviation

can be found between the pitches  $A\flat_3$  and  $A_3$ . Note that the audio distribution exhibits only one peak located between  $A\flat_3$  and  $A_3$ . A similar, but less salient deviation can be observed in the pitch range between  $D\flat_3$  and  $D_3$ .

In order to further investigate these discrepancies, we fit a Gaussian Mixture Model (GMM) with 13 Gaussians to the audio distribution from **Figure 7d**. The resulting mixture distribution is shown in **Figure 8**. The centers of the Gaussian pitch clusters are denoted with black numbers on top of the peaks, while the intervals between neighboring clusters are indicated with red numbers in between. The intervals between cluster centers from left to right are 192, 191, 152, 191, 173, 166, 155, 213, 178, 146, 179, and 196 cents. The numbers show that all intervals are all significantly larger than a semi-tone (100 cents) and most of them are smaller than a whole tone (200 cents). From these results, we can draw two conclusions: first, the sung intervals indicate—once more—that Artem Erkomaishvili’s tuning is clearly not equal-tempered. Second, melodic steps between 100 and 200 cents can sometimes be perceived and transcribed as minor 2nd, sometimes as major 2nd. As a consequence, this can lead to effects in the transcription like in **Figure 7a** (QNR 10–14), where  $A$ ,  $A\flat$ ,  $D$ , and  $D\flat$  appear closely together in time. From a Western (tempered) perspective, this might seem counter intuitive. However, this is merely an effect of forcing a not equal-tempered tuning system into tempered Western staff notation. The task gets even more challenging for the transcriber if additional constraints by the harmonic context are imposed (e.g., the harmonic fifth between bass and middle voice at QNR 13).

In summary, these case studies show the potential of our annotations for studies on Artem Erkomaishvili’s performances, as well as for analyzing tuning, pitch inventory, and musical scales underlying traditional Georgian vocal music.

## 7. Conclusions

In this paper, we presented a carefully organized, manually annotated, and publicly available dataset of traditional Georgian vocal music. The corpus is based on historic recordings of the former master chanter Artem Erkomaishvili. As part of our work, we collated existing

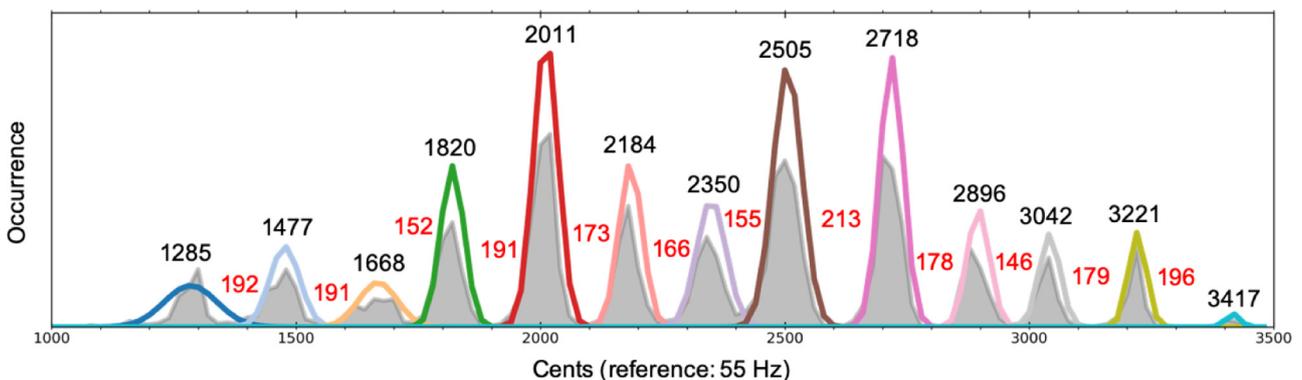
audio data and annotations. Furthermore, we generated onset annotations based on the digitized transcriptions by Shugliashvili (2014). Finally, we developed an interactive web-based user interface with score-following audio players, which provides convenient access to the corpus data. Beside contributing to the preservation and dissemination of the rich Georgian musical heritage, this dataset is a versatile resource for MIR research and empowers musicological research on traditional Georgian vocal music.

## Notes

- 1 Picture of Artem Erkomaishvili from (Shugliashvili, 2014).
- 2 <http://www.alazani.ge/old-archives-Artem-Erkomaishvilis-Sagaloblebi-folk-songs-ans59.html>.
- 3 <https://dunya.compmusic.upf.edu/>.
- 4 <http://liederenbank.nl/>.
- 5 <https://www.polyphonyproject.com>.
- 6 Further deviations are documented in the web-based interface (see Section 5).
- 7 <https://www.klemm-music.de/makemusic/finale/>.
- 8 <https://www.avid.com/de/sibelius>.
- 9 <https://zenodo.org/>.
- 10 <https://www.audiolabs-erlangen.de/resources/MIR/2019-GeorgianMusic-Erkomaishvili>.

## Acknowledgements

This work was supported by the German Research Foundation (DFG MU 2686/13-1, SCHE 280/20-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS. We thank Selina Sonntag, Stefanie Kämmerle and Moritz Berendes for helping us with the annotations. We gratefully acknowledge the support by the Polyphonies Vivantes Association for bringing the transcriptions of Shugliashvili (2014) into digital form and making them available for the current analysis. Furthermore, we would like to thank Frank Zalkow, Lukas Dietz, and El Mehdi Lemnaouar for their contributions to the web-based interface and Nana Mzhavanadze for her feedback regarding the representation of not equal-tempered music.



**Figure 8:** GMM with 13 Gaussians fitted to pitch histogram determined from adapted F0-trajectories of the song “Gushin Shentana” (GCH-ID 010). The black numbers on top of the peaks indicate the peak values in cents, while the red numbers indicate the intervals between neighboring peaks. The original distribution is indicated in grey color.

## Competing Interests

The authors have no competing interests to declare.

## References

- Arzt, A.** (2016). *Flexible and robust music tracking*. PhD thesis, Johannes Kepler Universität Linz.
- Böck, S., Krebs, F., & Schedl, M.** (2012). Evaluating the online capabilities of onset detection methods. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 49–54.
- Boersma, P.** (2001). Praat, a system for doing phonetics by computer. *Glott International*, 5(9/10), 341–345.
- Bogdanov, D., Wack, N., Gómez, E., Gulati, S., Herrera, P., Mayor, O., Roma, G., Salamon, J., Zapata, J. R., & Serra, X.** (2013). Essentia: An audio analysis library for music information retrieval. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 493–498. Curitiba, Brazil. DOI: <https://doi.org/10.1145/2502081.2502229>
- Byrd, D., & Simonsen, J. G.** (2015). Towards a standard testbed for optical music recognition: Definitions, metrics, and page images. *Journal of New Music Research*, 44(3), 169–195. DOI: <https://doi.org/10.1080/09298215.2015.1045424>
- Cannam, C., Landone, C., & Sandler, M. B.** (2010). Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the International Conference on Multimedia*, pages 1467–1468. Florence, Italy. DOI: <https://doi.org/10.1145/1873951.1874248>
- Cano, E., FitzGerald, D., Liutkus, A., Plumbley, M. D., & Stöter, F.** (2019). Musical source separation: An introduction. *IEEE Signal Processing Magazine*, 36(1), 31–40. DOI: <https://doi.org/10.1109/MSP.2018.2874719>
- Chokhanelidze, E.** (2010). Some characteristic features of the voice coordination and harmony in Georgian multipart singing. In *Echoes from Georgia: Seventeen Arguments on Georgian Polyphony*, pages 135–145. Nova Science Publishers.
- Cuesta, H., Gómez, E., Martorell, A., & Loáiciga, F.** (2018). Analysis of intonation in unison choir singing. In *Proceedings of the International Conference of Music Perception and Cognition (ICMPC)*, pages 125–130. Graz, Austria.
- 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 (ISMIR)*, pages 637–642. Utrecht, The Netherlands.
- Dzhambazov, G., Srinivasamurthy, A., Sentürk, S., & Serra, X.** (2016). On the use of note onsets for improved lyrics-to-audio alignment in Turkish makam music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 716–722. New York City, USA.
- Erkvanidze, M.** (2016). The Georgian musical system. In *Proceedings of the International Workshop on Folk Music Analysis*, pages 74–79. Dublin, Ireland.
- Ganguli, K. K., & Rao, P.** (2018). On the distributional representation of ragas: Experiments with allied raga pairs. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 1(1), 79–95. DOI: <https://doi.org/10.5334/tismir.11>
- Gasser, M., Arzt, A., Gadermaier, T., Grachten, M., & Widmer, G.** (2015). Classical music on the web – user interfaces and data representations. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 571–577. Málaga, Spain.
- Godsill, S., Rayner, P., & Cappé, O.** (2002). Digital audio restoration. In *Applications of Digital Signal Processing to Audio and Acoustics*, pages 133–194. Springer. DOI: [https://doi.org/10.1007/0-306-47042-X\\_4](https://doi.org/10.1007/0-306-47042-X_4)
- Gómez, E., Herrera, P., & Gómez-Martin, F.** (2013). Computational ethnomusicology: Perspectives and challenges. *Journal of New Music Research*, 42(2), 111–112. DOI: <https://doi.org/10.1080/09298215.2013.818038>
- Gong, R., Repetto, R. C., & Serra, X.** (2017). Creating an a cappella singing audio dataset for automatic jingju singing evaluation research. In *Proceedings of the International Workshop on Digital Libraries for Musicology*, pages 37–40. DOI: <https://doi.org/10.1145/3144749.3144757>
- Graham, J.** (2015). *The transcription and transmission of Georgian Liturgical Chant*. PhD thesis, Princeton University.
- Jeong, D., Kwon, T., Park, C., & Nam, J.** (2017). PerformScore: Toward performance visualization with the score on the web browser. In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China.
- Jgharkava, I.** (2016). Pearls of Georgian Chant. CD. Produced by the Georgian Chanting Foundation & Tbilisi State Conservatoire.
- Kroher, N., Díaz-Báñez, J. M., Mora, J., & Gómez, E.** (2016). Corpus COFLA: A research corpus for the computational study of flamenco music. *Journal on Computing and Cultural Heritage (JOCCH)*, 9(2), 10:1–10:21. DOI: <https://doi.org/10.1145/2875428>
- Lartillot, O., & Toivainen, P.** (2007). MIR in MATLAB (II): A toolbox for musical feature extraction from audio. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 127–130. Vienna, Austria.
- McFee, B., Raffel, C., Liang, D., Ellis, D. P., McVicar, M., Battenberg, E., & Nieto, O.** (2015). Librosa: Audio and music signal analysis in python. In *Proceedings of the Python Science Conference*, pages 18–25. DOI: <https://doi.org/10.25080/Majora-7b98e3ed-003>
- Müller, M., Arzt, A., Balke, S., Dorfer, M., & Widmer, G.** (2019). Cross-modal music retrieval and applications: An overview of key methodologies. *IEEE Signal Processing Magazine*, 36(1), 52–62. DOI: <https://doi.org/10.1109/MSP.2018.2868887>
- Müller, M., Grosche, P., & Wiering, F.** (2009). Robust segmentation and annotation of folk song recordings. In *Proceedings of the International Society for Music*

- Information Retrieval Conference (ISMIR)*, pages 735–740. Kobe, Japan.
- Müller, M., Rosenzweig, S., Driedger, J., & Scherbaum, F.** (2017). Interactive fundamental frequency estimation with applications to ethnomusicological research. In *Proceedings of the AES International Conference on Semantic Audio*, pages 186–193. Erlangen, Germany.
- Müller, M., & Zalkow, F.** (2019). FMP notebooks: Educational material for teaching and learning fundamentals of music processing. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, Delft, The Netherlands.
- Panteli, M.** (2018). *Computational analysis of world music corpora*. PhD thesis, Queen Mary University of London, UK.
- Pugin, L., Zitellini, R., & Roland, P.** (2014). Verovio: A library for engraving MEI music notation into SVG. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 107–112. Taipei, Taiwan.
- Rebello, A., Fujinaga, I., Paszkiewicz, F., Marcal, A. R. S., Guedes, C., & Cardoso, J. S.** (2012). Optical music recognition: State-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1(3), 173–190. DOI: <https://doi.org/10.1007/s13735-012-0004-6>
- Repetto, R. C., Pretto, N., Chaachoo, A., Bozkurt, B., & Serra, X.** (2018). An open corpus for the computational research of Arab-Andalusian music. In *Proceedings of the International Conference on Digital Libraries for Musicology*, pages 78–86. Paris, France. DOI: <https://doi.org/10.1145/3273024.3273025>
- Repetto, R. C., & Serra, X.** (2014). Creating a corpus of Jingju (Beijing Opera) music and possibilities for melodic analysis. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 313–318. Taipei, Taiwan.
- Rosenzweig, S.** (2017). Audio processing techniques for analyzing Georgian vocal music. Master Thesis, Friedrich-Alexander-Universität Erlangen-Nürnberg.
- Rosenzweig, S., Scherbaum, F., & Müller, M.** (2019). Detecting stable regions in frequency trajectories for tonal analysis of traditional Georgian vocal music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 352–359. Delft, The Netherlands.
- Röwenstrunk, D., Prätzlich, T., Betzwieser, T., Müller, M., Szwillus, G., & Veit, J.** (2015). Das Gesamtkunstwerk Oper aus Datensicht – Aspekte des Umgangs mit einer heterogenen Datenlage im BMBF-Projekt “Freischütz Digital”. *Datenbank-Spektrum*, 15(1), 65–72. DOI: <https://doi.org/10.1007/s13222-015-0179-0>
- Salamon, J., & Gómez, E.** (2012). Melody extraction from polyphonic music signals using pitch contour characteristics. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6), 1759–1770. DOI: <https://doi.org/10.1109/TASL.2012.2188515>
- Salamon, J., Gómez, E., Ellis, D. P. W., & Richard, G.** (2014). Melody extraction from polyphonic music signals: Approaches, applications, and challenges. *IEEE Signal Processing Magazine*, 31(2), 118–134. DOI: <https://doi.org/10.1109/MSP.2013.2271648>
- Scherbaum, F.** (2016). On the benefit of larynxmicrophone field recordings for the documentation and analysis of polyphonic vocal music. *Proceedings of the International Workshop Folk Music Analysis*, pages 80–87.
- Scherbaum, F., Müller, M., & Rosenzweig, S.** (2017). Analysis of the Tbilisi State Conservatory recordings of Artem Erkomaishvili in 1966. In *Proceedings of the International Workshop on Folk Music Analysis*, pages 29–36. Málaga, Spain.
- Scherbaum, F., Mzhavanadze, N., Rosenzweig, S., & Müller, M.** (2019). Multi-media recordings of traditional Georgian vocal music for computational analysis. In *Proceedings of the International Workshop on Folk Music Analysis*, pages 1–6. Birmingham, UK.
- Şentürk, S.** (2016). *Computational analysis of audio recordings and music scores for the description and discovery of Ottoman-Turkish Makam music*. PhD thesis, Universitat Pompeu Fabra.
- Serra, X.** (2014a). Computational approaches to the art music traditions of India and Turkey. *Journal of New Music Research, Special Issue on Computational Approaches to the Art Music Traditions of India and Turkey*, 43(1), 1–2. DOI: <https://doi.org/10.1080/09298215.2014.894083>
- Serra, X.** (2014b). Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the AES International Conference on Semantic Audio*, London, UK.
- Shugliashvili, D.** (2014). *Georgian Church Hymns, Shemokmedi School*. Georgian Chanting Foundation.
- Six, J., Cornelis, O., & Leman, M.** (2013). Tarsos, a modular platform for precise pitch analysis of Western and non-Western music. *Journal of New Music Research*, 42(2), 113–129. DOI: <https://doi.org/10.1080/09298215.2013.797999>
- Srinivasamurthy, A., Koduri, G. K., Gulati, S., Ishwar, V., & Serra, X.** (2014). Corpora for music information research in Indian art music. In *Proceedings of the Joint Conference 40th International Computer Music Conference (ICMC) and 11th Sound and Music Computing Conference (SMC)*, Athens, Greece.
- Thomas, V., Fremerey, C., Müller, M., & Clausen, M.** (2012). Linking sheet music and audio – challenges and new approaches. In Müller, M., Goto, M., & Schedl, M., Editors, *Multimodal Music Processing*, volume 3 of *Dagstuhl Follow-Ups*, pages 1–22. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany.
- Tsereteli, Z., & Veshapidze, L.** (2014). On the Georgian traditional scale. In *Proceedings of the International Symposium on Traditional Polyphony*, pages 288–295. Tbilisi, Georgia.
- Tzanetakis, G.** (2009). Music analysis, retrieval and synthesis of audio signals MARSYAS. In *Proceedings of the ACM International Conference on Multimedia (ACM-MM)*, pages 931–932. Vancouver, British Columbia, Canada. DOI: <https://doi.org/10.1145/1631272.1631459>

- Tzanetakis, G.** (2014). Computational ethnomusicology: A music information retrieval perspective. In *Proceedings of the Joint Conference 40th International Computer Music Conference (ICMC) and 11th Sound and Music Computing Conference (SMC)*, pages 69–73. Athens, Greece.
- Tzanetakis, G., Kapur, A., Schloss, W. A., & Wright, M.** (2007). Computational ethnomusicology. *Journal of Interdisciplinary Music Studies*, 1(2), 1–24.
- Uyar, B., Atli, H. S., Sentürk, S., Bozkurt, B., & Serra, X.** (2014). A corpus for computational research of Turkish makam music. In *Proceedings of the International Workshop on Digital Libraries for Musicology*, pages 1–7. London, UK. DOI: <https://doi.org/10.1145/2660168.2660174>
- van Kranenburg, P., de Bruin, M., & Volk, A.** (2019). Documenting a song culture: The Dutch Song Database as a resource for musicological research. *International Journal on Digital Libraries*, 20(1), 13–23. DOI: <https://doi.org/10.1007/s00799-017-0228-4>
- Werner, N., Balke, S., Stöter, F.-R., Müller, M., & Edler, B.** (2017). trackswitch.js: A versatile webbased audio player for presenting scientific results. In *Proceedings of the Web Audio Conference (WAC)*, London, UK.
- Zalkow, F., Rosenzweig, S., Graulich, J., Dietz, L., Lemnaouar, E. M., & Müller, M.** (2018). A web-based interface for score following and track switching in choral music. In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France.

**How to cite this article:** Rosenzweig, S., Scherbaum, F., Shugliashvili, D., Arifi-Müller, V., & Müller, M. (2020). Erkomaishvili Dataset: A Curated Corpus of Traditional Georgian Vocal Music for Computational Musicology. *Transactions of the International Society for Music Information Retrieval*, 3(1), pp. 31–41. DOI: <https://doi.org/10.5334/tismir.44>

**Submitted:** 21 October 2019

**Accepted:** 03 February 2020

**Published:** 10 April 2020

**Copyright:** © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.

**OPEN ACCESS** 