MUSICOLOGIST
International Journal of Music Studies
Trabzon University State Conservatory © 2017-2022
Volume **6** Issue **2** **December 2022**

Research Article
Musicologist 2022. 6 (2): 142-168
DOI: 10.33906/musicologist.1068947

FRANK SCHERBAUM
Universität Potsdam, Germany
fs@geo.uni-potsdam.de
orcid.org/0000-0002-5050-7331

NANA MZHAVANADZE
State Folk Centre, Georgia
nanateknada@gmail.com
orcid.org/0000-0001-5726-1656

SEBASTIAN ROSENZWEIG
International Audio Laboratories Erlangen, Germany
sebastian.rosenzweig@audiolabs-erlangen.de
orcid.org/0000-0003-4964-9217

MEINARD MÜLLER
International Audio Laboratories Erlangen, Germany
meinard.mueller@audiolabs-erlangen.de
orcid.org/0000-0001-6062-7524

Tuning Systems of Traditional Georgian Singing Determined from a New Corpus of Field Recordings

ABSTRACT
In this study we examine the tonal organization of the 2016 GVM dataset, a newly-created corpus of high-quality multimedia field recordings of traditional Georgian singing with a focus on Svaneti. For this purpose, we developed a new processing pipeline for the computational analysis of non-western polyphonic music which was subsequently applied to the complete 2016 GVM dataset. To evaluate under what conditions a single tuning system is representative of current Svan performance practice, we examined the stability of the obtained tuning systems from an ensemble-, a song-, and a corpus-related perspective. Furthermore, we compared the resulting Svan tuning systems with the tuning systems obtained for the Erkomaishvili dataset (Rosenzweig et al., 2020) in the study by Scherbaum et al. (2020). In comparison to a 12-TET (12-tone-equal- temperament) system, the Erkomaishvili and the Svan tuning systems are surprisingly similar. Both systems show a strong presence of pure fourths (500 cents) and fifths (700 cents), and 'neutral' thirds (peaking around 350 cents) as well as 'neutral' sixths. In addition, the sizes of the melodic and the harmonic seconds in both tuning systems differ systematically from each other, with the size of the harmonic second being systematically larger than the melodic one.

KEYWORDS
Traditional Georgian Music
Tuning
Computational ethnomusicology

Received: February 6, 2022; **Accepted:** September 1, 2022

Introduction

The rich musical heritage of the country of Georgia has attracted the attention of musicians, music lovers and music scholars for a long time. Over the years, starting with phonograph recordings already more than a century ago, many efforts have been made to record and document traditional Georgian music. However, many of these recordings have been lost in the course of time and the available historic audio tracks are often of insufficient quality for the application of modern, quantitative analysis techniques. One well-known exception is the collection of the Tbilisi State Conservatory recordings of master chanter Artem Erkomaishvili from the year 1966, the computational analysis of which is covered in a series of papers (Müller et al., 2017; Scherbaum et al., 2017; 2020; Rosenzweig et al., 2019; 2020; 2021). In preparation for the research project "Computational Analysis of Traditional Georgian Vocal Music" (GVM)[1], two of the authors (N.M. and F.S.) travelled through Georgia during summer of 2016 to record a new research corpus of traditional Georgian singing, praying, and lamenting. All of the recordings were done as multichannel recordings in which a high resolution (4K) video stream was combined with a stream of 3-channel headset microphone recordings (one for each voice group), a stream of 3-channel larynx microphone recordings (also one for each voice group), and a conventional stereo recording. The use of larynx microphones, which had never before been systematically employed in ethnomusicological field work, allows the use of methods from audio signal processing and music information retrieval (MIR) to analyse the recordings of the separate voices while the singers are singing together in their natural singing environment.

The regional focus of the field expedition was on Upper Svaneti which is one of the rare regions at the crossroads of Europe and Asia where very old (presumably pre-Christian) traditions are still cultivated as part of daily life. Svan songs as parts of these rituals therefore occupy a special place within Georgian music and are still maintained in a comparatively original form on account of the remote geographical situation. The reason for choosing Svaneti as our main target was that presumably the first stages of Georgian vocal music development (and possibly of Europe) have been preserved there (Jordania,

[1] Computational Analysis of Traditional Georgian Vocal Music (GVM) [DFG MU 2686/13-1, SCHE 280/20-1] (2018 - 2022)

2006).

The complete dataset, the technical details of which are described in Scherbaum et al. (2018) and Scherbaum et al. (2019), has been made publicly available on two websites[2]. In addition to the systematic collection of recordings from Upper Svaneti, it contains recordings of two ensembles from Guria, a small collection of recordings of singers from Racha, and a set of recordings of a Tbilisi-based ensemble.

In recent years, we have witnessed a revolution in the way computer technology has affected the way we do research, not only in the natural sciences. Not surprisingly, these changes have also impacted ethnomusicology and led to the emergence of a new field of research called "computational ethnomusicology" (see Tzanetakis et al., 2007; Gómez et al., 2013; Tzanetakis, 2014). The intellectual core of computational ethnomusicology is what the influential mathematician Stephen Wolfram has called "computational thinking, which he describes as the principle of " formulating questions with enough clarity, and in a systematic enough way, that one can tell a computer how to do them" (Wolfram, 2016). Of course, not all relevant research questions in the field of ethnomusicology can be solved computationally, but the analysis of audio signals and the extraction of quantitative information from them are prime examples of where computational approaches are extremely useful.

The starting points for the present study are computationally estimated pitch trajectories for the individual voices of all recordings of the GVM 2016 dataset belonging to three-voice songs of sufficiently high fidelity for computational analysis. The technical goal of our analysis was the derivation of quantitative models for the melodic pitch distributions and harmonic interval distributions for the complete dataset, but also separately for the Svanetian, Gurian, Rachan, and the Tbilisi subsets. Thereby we aimed to investigate what these recordings can contribute to the ongoing scientific dispute about traditional Georgian tuning systems, a comprehensive review of which can be found in Section 2 of Scherbaum et al. (2020).

The rest of the current paper is structured as follows. After this introduction, we discuss the processing workflow to obtain the melodic pitch and harmonic interval distributions.

[2] https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Scherbaum and https://lazardb.gbv.de/search

In order to cope with the peculiarities of the singing styles in Svaneti and Guria (continuous pitch shifting and rapidly changing note durations, respectively), we had to develop and test new computational tools for music and audio processing and implement them into a completely new processing workflow. Then, in the results section, we present the pitch and interval distributions as so-called Gaussian mixture distributions, a very efficient numerical representation already used in Scherbaum et al. (2020), which also facilitates the comparison of the tuning systems of the different regional subsets. The final section of the paper contains a discussion and concluding remarks.

Processing Workflow

The first step of our processing workflow to derive pitch distributions and harmonic interval distributions, respectively, consists of processing each recorded audio track of interest (in our case preferably the output signals of the larynx microphones) with a so-called 'pitch[3]' tracking algorithm. For the present study we used the autocorrelation-based PYIN algorithm (Mauch and Dixon, 2014), which is also implemented in the widely used Tony software (Mauch et al., 2015). The resulting pitch trajectories, which are commonly also referred to as 'fundamental frequency' or 'F0' trajectories, were subsequently checked for octave jumps and artefacts. Octave jumps were corrected and obvious artefacts (e.g., non-singing related utterances such as clearing the throat, coughing, and so on) removed. For the analysis of pitch distributions and harmonic interval distributions, we are only interested in stable segments of the F0 trajectories (where one can perceive stable pitches) and not in utterances corresponding to very short transient signals such as sliding phases ('portamento'), which are a common feature, especially in Svan singing. A computationally efficient way to remove the latter is by morphological filtering, as described by Rosenzweig et al. (2019). Here we use an extension of the original algorithm of Rosenzweig et al. (2019), motivated by the fact that, especially for the Gurian songs in the 2016 GVM dataset, we were confronted with notes

[3] In the field of psychoacoustics, the term 'pitch' is defined as a quantity that can not be measured directly, but which is perceptually more closely related to an autocorrelation process than to the measurement of a physical frequency component (Heller, 2012). In contrast, in the field of acoustic sound analysis the term 'pitch' is usually understood as a numerical quantity which can be determined by means of so-called pitch tracking algorithms and which is also called 'F0' or 'fundamental frequency'. The latter term has to be taken with a grain of salt, since it is commonly also used for algorithms in which F0 is determined purely in the time domain, e.g. by using an autocorrelation-based approach which is the current state-of-the-art of pitch determination (Heller, 2012). Those algorithms will return proxies for 'pitch' values even for signals without spectral energy at F0, or for periodic non-harmonic signals, for which fundamental frequencies in the strict sense are not defined.

with rapidly varying durations that the original algorithm cannot handle optimally.

The extended approach first processes the raw pitch trajectories with the original algorithm, but uses extremely short filter lengths (e.g. three to five samples). This results in a large number of very short stable segments which we call 'note fragments'. These note fragments are subsequently recombined to new 'note objects' according to two criteria (motivated by simple perceptional principles).

The post-processing starts by taking the first note fragment as the start of a new (longer) note object. The following note fragment is joined with the previous one if a) the pitch and time differences between the two fall below chosen thresholds and b) if the pitch of the candidate note fragment to be joined stays within a chosen distance from the cumulative mean pitch value of all the previous fragments already joined in the long note object. This prevents the creation of note objects with excessively curved pitch curves. As an additional optional feature, we have also implemented the possibility of trimming the beginnings and ends of the newly generated note objects so that the pitch trajectory within it stays within a chosen maximum distance from its mean pitch value[4]. The performance of this processing scheme is illustrated in Fig. 1, which shows the first 20 seconds of the pitch trajectory of the top voice of the Gurian song "Chven Mshvidoba."[5]

[4] This was to facilitate the visual comparison with Tony-generated note objects displayed in Tony by blue rectangles.

[5] The corresponding audio signal can be accessed at https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Scherbaum at GVMID 017.

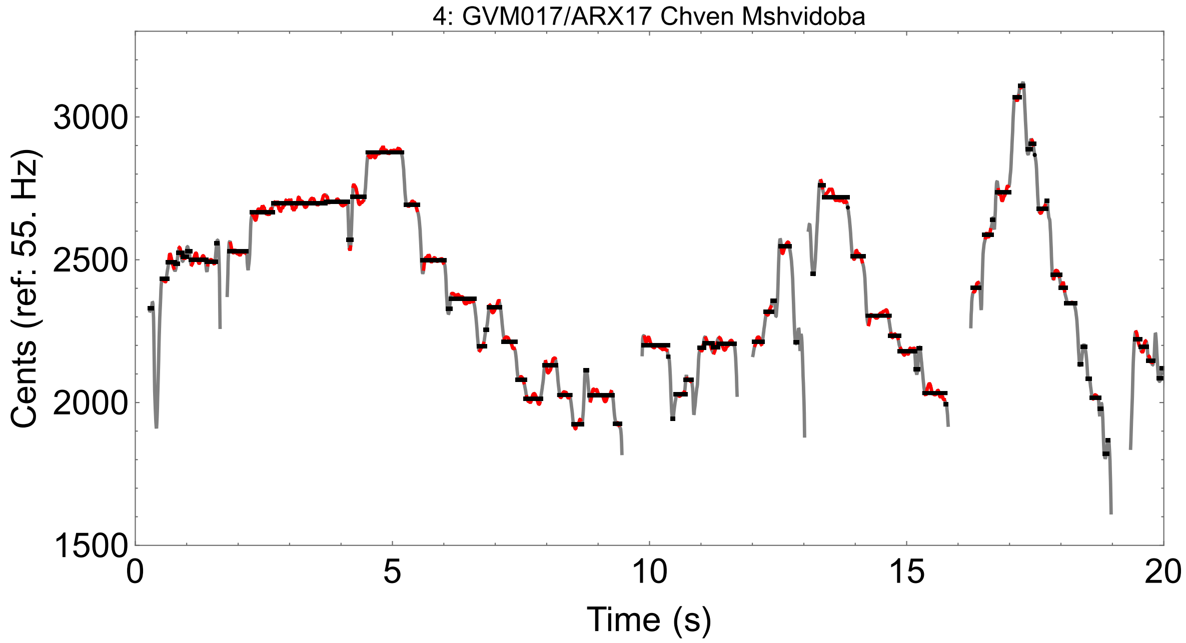


Figure 1. Determination of stable segments (note objects) of the pitch trajectory of the beginning of the top voice of the song 'Chven Mshvidoba'. The wiggly grey line corresponds to the raw pitch trajectory, the horizontal black lines to the detected note objects, and the wiggly red lines to the pitch trajectories within those note objects.

The black horizontal solid lines in Fig. 1 show the resulting stable segments for the beginning of the song "Chven Mshvidoba," sung by the trio Shalva Chemo (GVMID 017). The wiggly red lines show the corresponding parts of the pitch trajectory (after restriction to the stable segments). In principle, a similar result could also have been achieved with the software tool "Tony" (Mauch et al. 2015), which has the additional advantage of visually and acoustically controlling precisely which parts of an audio signal are used for the determination of note events. However, for the processing of the complete GVM dataset with several hundred tracks, this would have been far too time-consuming.

The greatest technical challenge which we were confronted with in the course of the data processing of the GVM dataset was caused by the fact that a large portion of the performances by Svan singers was accompanied by a gradual and coordinated rise of the singers' pitches of up to several hundred cents over the course of the whole song. Fig. 2. shows several examples to illustrate this phenomenon. Gradual pitch rises have been described by Georgian scholars (e.g. Paliashvili, Arakishvili, Garakanidze) for a long time, both for individual and for antiphonal performances. They are also well known from other unaccompanied vocal performance traditions worldwide. Specific examples of this phenomenon are discussed, for example, in Chapter 7 in Ambrazevičius et al., (2015).

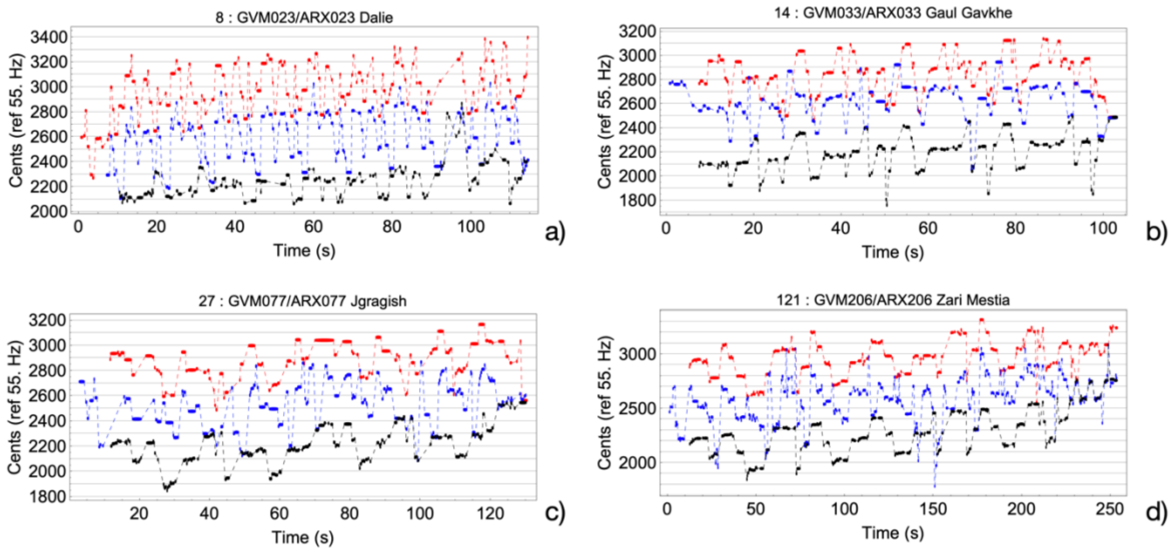


Figure 2. Display of the stable segments (note objects) of the pitch trajectories for a selected set of Svan song showing a steady pitch rise.

In the context of discussing the properties of eleven *Zär* recordings from the 2016 GVM dataset in a series of previous papers (Mzhavanadze and Scherbaum, 2020; Scherbaum and Mzhavanadze, 2020), a very labour-intensive annotation process was used to remove the gradual pitch rises for subsequent processing. This would not have been applicable to the entire 2016 GVM dataset on account of the time required. A recent study by Rosenzweig et al. (2022) demonstrated that the use of 'interval filtering' could significantly simplify and hence also speed up the removal of pitch drifts from larger datasets. The basic idea behind this approach is recognizing that Georgian singers employ what they sometimes refer to as 'vertical thinking': they mutually adjust the intonation of their voices to generate particular harmonic (vertical) intervals which are as 'pure' as possible. In other words, they may sacrifice the precision of a melodic line in exchange for an increase of precision of a harmonic interval. This in turn means that interval-filtered trajectories, leaving only those note pairs where the corresponding harmonic intervals meet chosen precision levels, will correspond to those locations in the time-pitch space where the singers obviously want to be and, consequently, will 'mark' the pitch rise. For Svan singing, the two prime interval candidates to be used for interval filtering are the unison and the fifth (702 cents).[6] These considerations led to a new efficient processing strategy for pitch rise removal (when combined with an interactive interface). Fig. 3 illustrates the related processing scheme for the song "Jgragish" (GVMID 077), for which the raw note trajectory is displayed in Fig. 2c.

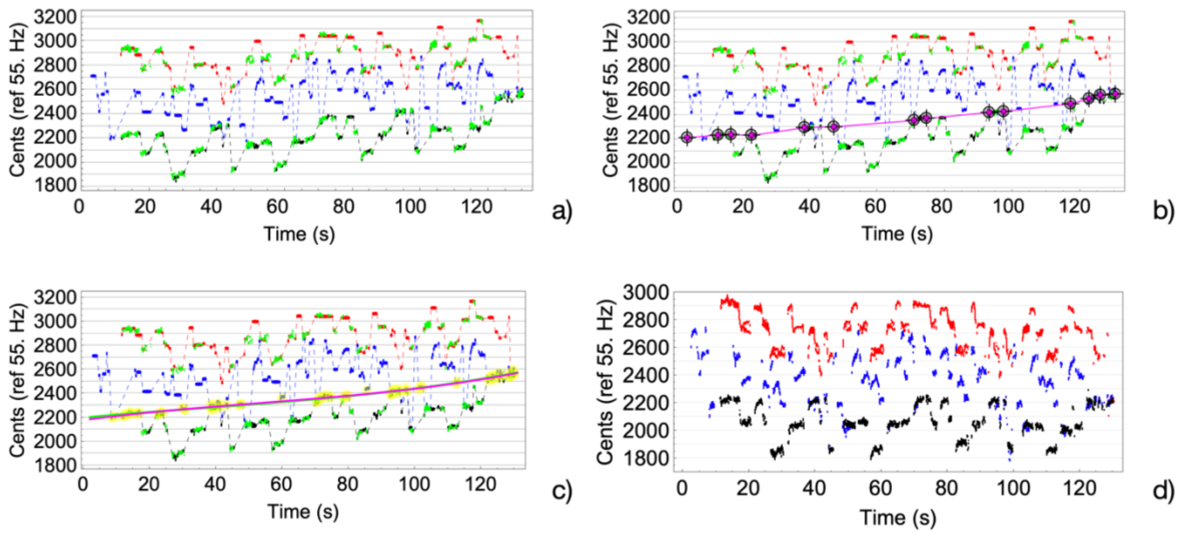


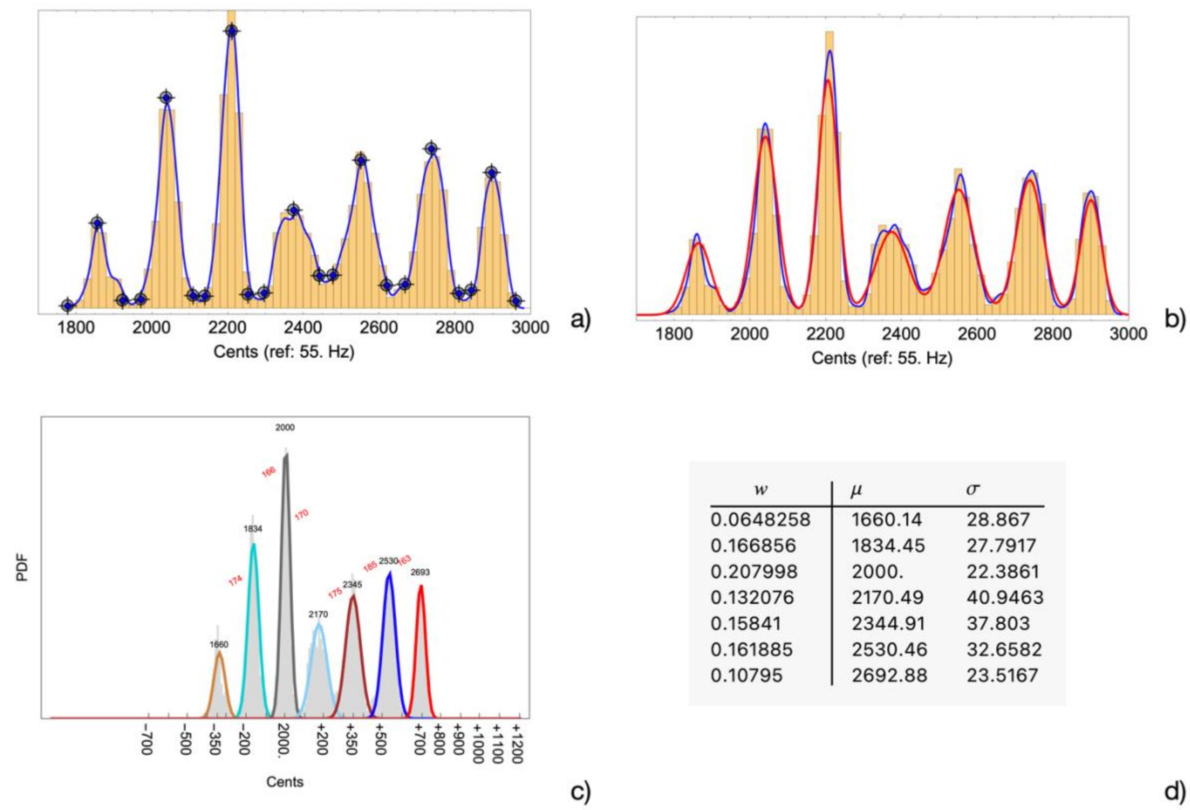
Figure 3. Illustration of the processing steps to identify and remove the steady pitch rise from the pitch trajectories of song "Jgragish," shown in Fig. 2c. We refer to the text for further explanations.

The green note objects in Fig. 3a show interval-filtered note trajectories for 0 ± 40 cents (unisons) and for 700 ± 40 cents (fifths). Fig. 3b illustrates the placement of locator objects (circles with crosses) to a selected scale degree of one of the voices, which is assumed to represent the gradual pitch rise. These placements were done manually using an interactive user interface. In this context, the display of the interval-filtered traces turned out to be very helpful. After placement, the position of the locator objects can still be moved until the analyst is satisfied. Once the locator placement has been accepted, two regression curves, displayed in green and magenta respectively in Fig. 3c, are calculated. The green curve corresponds to a third degree polynomial regression through the locator positions. The magenta curve corresponds to a third degree polynomial regression through all note objects found within a pitch differences ± 40 cents from the green regression curve. This puts more weight on the actual note objects and reduces the subjective influence of the analyst. Finally, after visual control, the magenta regression

[6] The octave, because of the close perceptional proximity of the two defining intervals, does not appear frequently enough in Svan songs to be useful.

curve is chosen to correct the pitches in the original note and pitch trajectories, the result of which is then displayed in Fig. 3d). This procedure turned out to be sufficiently efficient to be applicable to the whole 2016 GVM dataset.

The goal of the subsequent processing step is to find a compact representation for the pitch histograms of the drift-corrected pitch samples (from within the stable segments) of the pitch trajectories for all three voices of a song. These histograms typically show a strongly clustered structure with separated pitch clusters (at least if the pitch drift could be properly compensated for). Each of the pitch clusters can be seen to represent an individual melodic sound scale degree. The sand-coloured histogram in Fig. 4a shows the pitch histogram calculated from the complete set of pitches displayed in Fig. 3d.



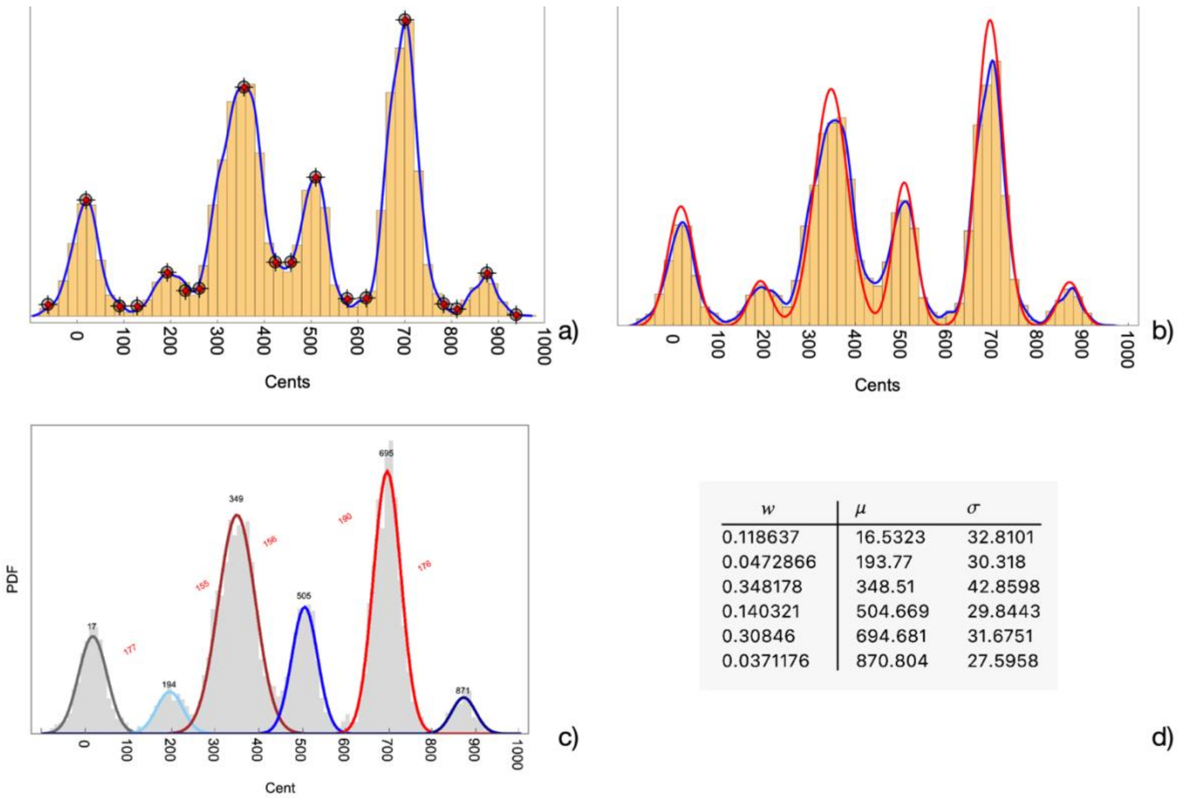
w	μ	σ
0.0648258	1660.14	28.867
0.166856	1834.45	27.7917
0.207998	2000.	22.3861
0.132076	2170.49	40.9463
0.15841	2344.91	37.803
0.161885	2530.46	32.6582
0.10795	2692.88	23.5167

Figure 4. Illustration of the determination of the Gaussian mixture model of the pitch distribution of the song "Jgragish" (GVMID 077). The black numbers on top of the pitch clusters show the μ_k-values of the corresponding mixture elements, while the tilted red numbers between two pitch groups correspond to the corresponding intervals in cents.

If one thinks of modelling each pitch cluster by a Gaussian distribution $\mathcal{N}(\mu, \sigma^2)$ with a mean value μ and a standard deviation σ, the complete pitch histogram can be represented as a so-called Gaussian Mixture Model (GMM). This is simply a weighted sum of individual Gaussian distributions $\mathcal{N}(\mu, \sigma^2)$. In the case of K pitch groups, this results in a representation as $\sum_{k=1}^{K} w_k\, \mathcal{N}(\mu_k, \sigma_k^2)$. The mean values of the individual Gaussians (the μ_k), which correspond to the centre values of the individual pitch clusters in the pitch histogram, can now be seen to express the pitches of the associated melodic sound scale degrees, while the standard deviations of the Gaussians (the σ_k) define the pitch variability within the associated scale degree. The w_k represent the individual weighting factors, in other words, how much a particular sound scale degree is present in a set of pitch trajectories.

Technically, the determination of a Gaussian mixture model is an iterative process, which requires starting values for the μ_k, the σ_k, and the w_k. Motivated by the efficiency of the interactive pitch drift correction, we implemented the determination of these starting values as another interactive procedure, which is illustrated in Fig. 4. In addition to the sand-coloured pitch histogram, Fig. 4a shows a blue solid line, resulting from smoothing the pitch histogram data (mathematically speaking representing it as a smooth kernel density distribution). The superimposed locators should be interpreted as triplets represeenting the pitch cluster's left boundary, peak, and right boundary, respectively. The locators marking the peaks are calculated from the smooth pitch histogram using a peak-finding algorithm, while the locators marking the left and right margins of a pitch cluster have to be placed manually by the analyst using the mouse cursor. After placement, the position of the locator objects can still be moved until the analyst is satisfied. Once the analysist accepts the locator placement, the position of the locator triplets is used to calculate the mean value and standard deviations for all the pitch values falling within the two locator margins. The peak value locator is used to calculate the starting value for the weighting factor. The resulting model, displayed in red in Fig. 4b, is used as initialization for estimating the Gaussian mixture distribution. Now it takes only very few iterations for the expectation maximization algorithm to converge to a final model which is shown in Fig. 4c. Furthermore, Fig. 4d shows the corresponding values for the w_k. the μ_k, and the σ_k for each Gaussian component in a row-wise fashion. In addition, prior to the generation of Fig. 4c, the pitch value for the final bass note was calculated and the complete distribution was adjusted in pitch so that the mean value of the pitch cluster containing the final bass note becomes 2000 cents so that the pitch distributions from different songs can easily be compared visually.

It seems worth mentioning for a non-technically inclined audience that this representation of the pitch distribution is not an end in itself but serves an essential purpose for subsequent analysis. This way, the complete set of pitches, the size of which for all three voices of a song of five-minute duration can easily exceed 100,000 pitch samples (in practice often one for every 5.8 milliseconds), the complete information for the subsequent task to analyse melodic sound scales, can be expressed by roughly 24 values per octave (for a sound scale with 8 scale degrees or pitch clusters per octave). Only with this kind of data reduction/compression does it become feasible at all to try to quantitatively compare hundreds of sound scale models derived from audio tracks and still be able to interpret the results.



w	μ	σ
0.118637	16.5323	32.8101
0.0472866	193.77	30.318
0.348178	348.51	42.8598
0.140321	504.669	29.8443
0.30846	694.681	31.6751
0.0371176	870.804	27.5958

Figure5. Illustration of the determination of the Gaussian mixture model for the harmonic interval distribution of the song "Jgragish" (GVMID 077). The black numbers on top of the pitch clusters show the μ_k-values of the corresponding mixture components while the tilted red numbers between two interval groups correspond to the differences between them. In order better to distinguish the individual Gaussians visually, they were displayed in different colours.

The last processing step consists of the calculation of the harmonic interval distribution. Technically, this is more or less identical to the processing of the pitch distributions except that it is performed on the samples of harmonic intervals, which are calculated from all the concomitant samples of the three voices. The resulting Gaussian mixture model for the harmonic interval distribution of the song "Jgragish" (GVMID 077) is shown in Fig. 5.

Results

The processing workflow described in the last section was applied to all those recordings in the GVM dataset that contained three simultaneously recorded voices and (based on visual inspection of the pitch trajectories) seemed of sufficient quality for processing. This left 125 potential performances, six of which had to be discarded for various technical reasons discovered later (e.g., the lack of sufficient numbers of concomitant samples, problems with time synchronization of channels, and so on). The remaining 119 recordings which were finally used in the present study correspond to 82 different songs from 19 different ensembles (14 from Svaneti, two from Racha, two from Guria, and one from Tbilisi). Since the choice of the performed songs was left to the ensembles, there is only partial overlap between the songs performed by the different groups. The maximum number of useable multiple recordings for a single song is five (for the song "Jgragish"). In the following subsections, we are going to address the following specific questions:

- How variable are the tuning systems used by ensembles?
- How similar are the specific tuning systems used by different ensembles for the same song?
- How similar are the average tuning systems used by different ensembles?

Ensemble Related Tuning Systems

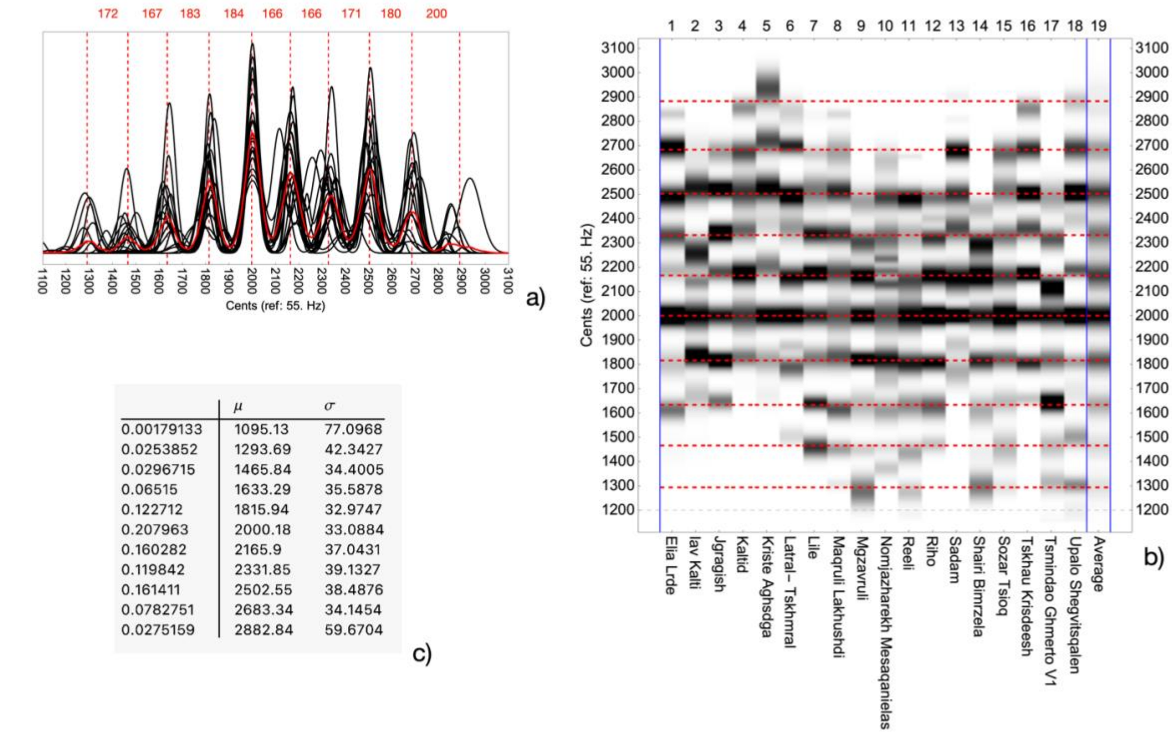
To determine the variability of ensemble-related tuning systems, we have used 18 recordings from the group Lakhushdi-B[7], a group of three male village singers from the village of Lakhushdi in Upper Svaneti. Since this is the largest number of recordings from a single ensemble, we consider this set the best one to address this question with some confidence. The next largest number of recordings from a single ensemble from Svaneti

[7] On purpose, we refer to all ensembles by a fictitious name, referring only to the location of origin of the singers.

is already down to nine, for which less stable variability estimates are to be expected (simply because of the smaller sample size).

The black solid curves in Fig 6a show the individual pitch distributions (as Gaussian mixture models) for all 18 songs. All of them were processed in the way described in the previous section. The solid red line shows the corresponding 'average pitch distribution'. The latter was calculated by first resampling the individual Gaussian mixture distributions for each song with the same number of samples (1,000). This was done to avoid a bias towards individual songs (which are of different durations and therefore contain different numbers of samples). Subsequently, a new Gaussian mixture model was calculated from the combination of samples. The resulting weights, mean values, and standard deviations are listed in Fig. 6c. The corresponding pitch distribution is shown in Fig 6a by the solid red curve. As expected, it maps the characteristic structure of the set of individual pitch distributions in an average sense. The red numbers on top of the panel in Fig. 6a denote the intervals between the centre values of the individual pitch groups (of the average distribution) in cents. Note that the pitch trajectories of all songs were adjusted so that the pitch of the cluster which contained the final note (the 'finalis') receives a value of 2000 cents. This processing step is a prerequisite for the comparison of different songs since all of them were sung in different absolute pitches. In order to be able to visually to compare and identify the characteristic features of the pitch distributions for the individual songs, Fig. 6b shows them as a density grid. Each of the 19 columns corresponds to the pitch distribution of one song the name of which is given at the bottom. The number at the top corresponds to the song index. Column 19 corresponds to the average model shown in red in Fig. 6a. Light grey tones correspond to small, and dark grey tones to larger amplitudes. The dashed red lines in Fig. 6a and Fig. 6b show the mean values of the Gaussian mixture model for the average distribution, which one could interpret as scale degree pitches.

For the average distribution, the pitch difference between the pitch of the finalis (2,000 cents) and the third and fourth pitch group above are 503 cents (pure melodic fourth) and 683 cents (slightly flat fifth), respectively. Interesting to note in this context is the observation that the pitch differences between neighbouring pitch groups and except for one are all significantly below 200 cents[8]. Therefore, even at first glance, it becomes obvious that the tuning system used by the group Lakhushdi-B deviates significantly from a 12-TET (12-tone-equal-temperament) system. This becomes even more obvious for the harmonic interval distributions shown in Fig. 7. The solid red line in Fig. 7a shows (again) the average distribution calculated from all 18 recordings. In contrast to the melodic tuning system, the harmonic fourth and fifth are amazingly pure with 504 and 696 cents, respectively. This is not surprising, since Svan music has a wealth of homophonic chord sequences with fourths and fifths, underlining the important role of harmonic thinking in Svan music.



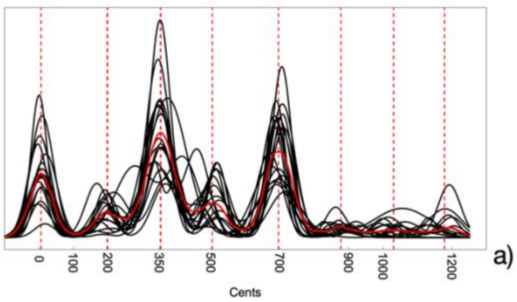
	μ	σ
0.00179133	1095.13	77.0968
0.0253852	1293.69	42.3427
0.0296715	1465.84	34.4005
0.06515	1633.29	35.5878
0.122712	1815.94	32.9747
0.207963	2000.18	33.0884
0.160282	2165.9	37.0431
0.119842	2331.85	39.1327
0.161411	2502.55	38.4876
0.0782751	2683.34	34.1454
0.0275159	2882.84	59.6704

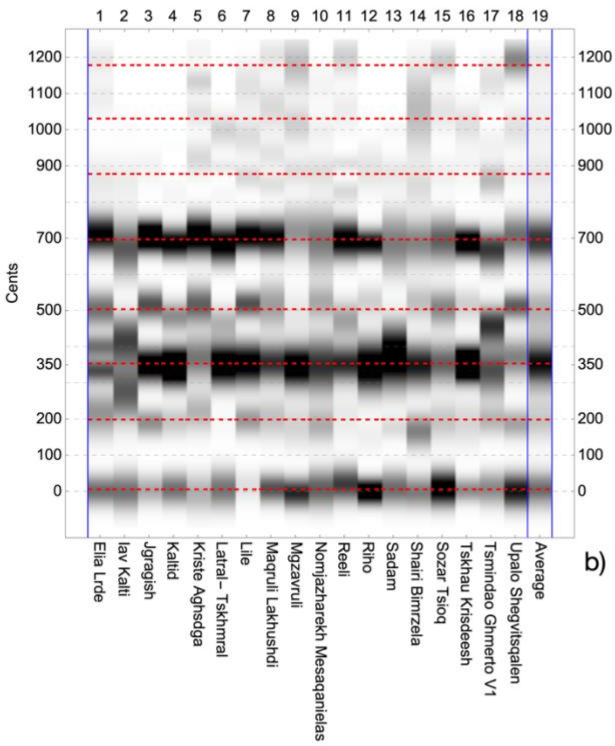
Figure 6. Pitch distributions for all 18 performances recorded with the ensemble Lakhushdi-B. For details see the text.

Two other features of the harmonic interval distributions are worth mentioning. First, the interval group corresponding to the harmonic thirds has a peak at the value of 353 cents (a ‘neutral’ third). With the exception of the second song ("Iav Kalti"), minor and major harmonic thirds are missing completely. The corresponding second column in Fig.

[8] The value of 200 cents is probably affected by a seeming ‘outlier group’ which peaks between 2900 and 3000 cents.

7b, however, shows two dark peaks at 300 and 400 cents, respectively, which can also be identified in Fig. 7a. Second, another noticeable exception from the otherwise very simple structure of the collection of harmonic interval distributions in Fig. 7b is related to the 17. song ("Tsmindao Ghmerto V1"). In this case, all the harmonic intervals (except the harmonic second at 200 cents) were sung significantly flat in comparison to the performances of the rest of the 18 songs. Revisiting the pitch- and note trajectories (not shown in this paper) revealed that in this particular case, the middle and bass voice singers consistently sang smaller harmonic intervals for this song, (which is the only Svan church song sung with Georgian text). In contrast, the top voice singer kept his harmonic major second to the middle voice singer consistently throughout the song. This is an interesting observation by itself. It tells us that the top voice singer oriented himself harmonically closer to his nearest neighbour, the middle voice singer with whom he sang a 'dissonant' major second than to the bass voice singer with whom he could have chosen to sing a 'consonant' fifth. In other words, he seemingly preferred the 'dissonant' over the 'consonant' interval.





	μ	σ
0.153208	5.66841	33.6274
0.0724311	198.709	38.7112
0.326297	353.542	42.6102
0.102604	503.615	38.9746
0.256636	696.433	40.1348
0.032167	877.598	41.2568
0.0291197	1030.56	49.6551
0.0275377	1178.42	36.5655

c)

Figure 7. Harmonic interval distributions for all 18 performances recorded with the group Lakhushdi-B. For details see the text.

To wrap up the discussion regarding the ensemble related tuning systems for the group Lakhushdi-B, we make the assumption that it is justified to use the average pitch and harmonic interval distributions as a proxy for the characteristic tuning of an ensemble. However, this assumption should not be made in a strict sense. Instead, one should keep in mind that the tuning systems used in individual songs might differ from the average significantly. Judging from the standard deviations for the pitch and interval groups in Fig. 6 b and Fig. 7b the corresponding precision for the respective means of the average models is on the order of ±40 cents.

Song Related Tuning Systems

In the following, we address the question of whether different ensembles use a similar tuning system when performing the same song. The most appropriate subset for the investigation of this topic is a set of five recordings of the song "Jgragish." The resulting pitch and harmonic interval distributions are displayed in Fig. 8. The left panels correspond to the pitch and the right panels to the harmonic interval distributions, respectively.

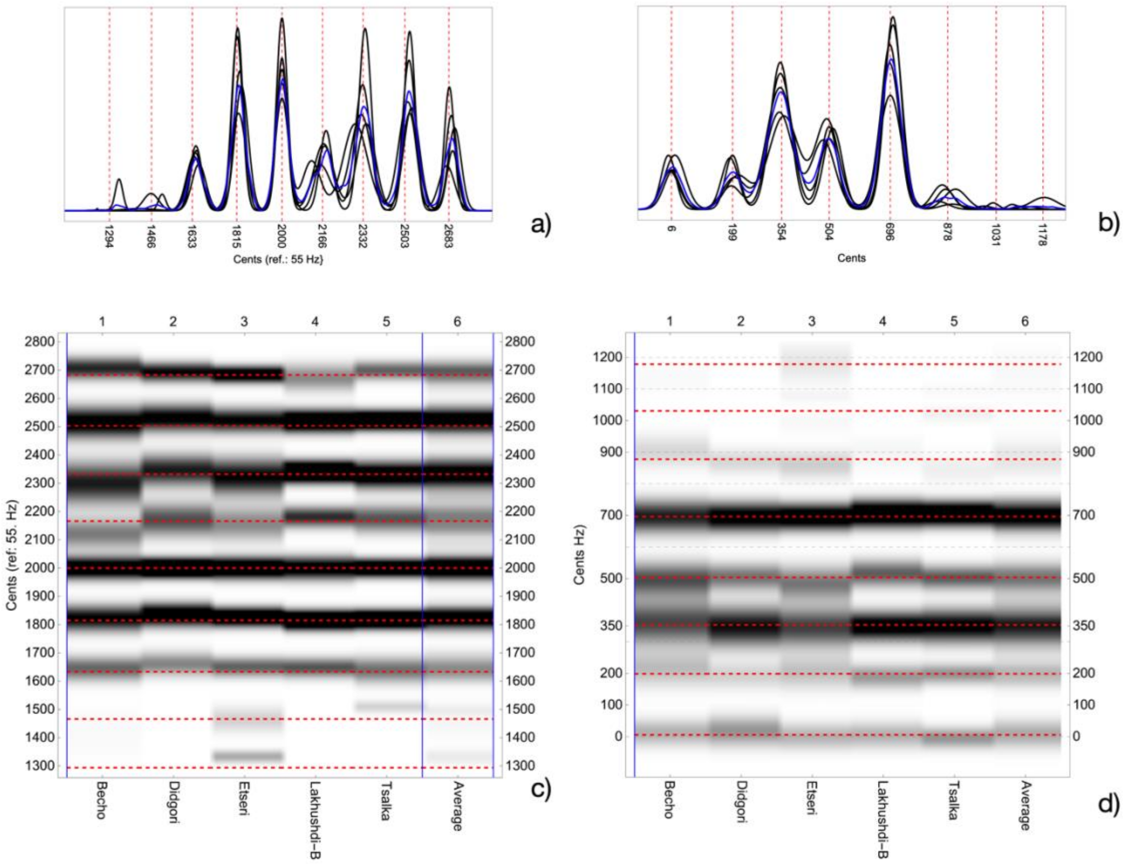


Figure 8. Pitch (left) and harmonic (right) interval distributions for the recordings of the song "Jgragish" performed by five different ensembles along with the average distributions.

Judging from the high degree of similarity of both distributions, we conclude that all five ensembles have a similar melodic and harmonic understanding of the song. However, one can also see in panels Fig. 8c and Fig. 8d that the bass voices, which contribute primarily to the pitch values below the finalis (2000 cents) differ between the groups Becho, Didgori, and Lakhushdi-B, on the one hand, and Etseri and Tsalka, which also differ between each other, on the other hand[9].

Comparison of All Ensemble-Related Distributions

In this section, we compare all the average tuning systems used by the different ensembles included in the 2016 GVM dataset. The recording locations are shown in Fig. 9. The recordings of the non-Svan ensembles are included only for completeness (with respect to the 2016 GVM dataset), yielding a superficial comparison. The small number

[9] The bass singers from Etseri for example use a pitch range down to approx. 1300 cents which the other ensembles did not use.

of ensembles from Guria, Racha, and Tbilisi does not allow an in-depth analysis similar to what we try to achieve for the Svan subset, for which a total of 83 performances from 14 different ensembles is available.

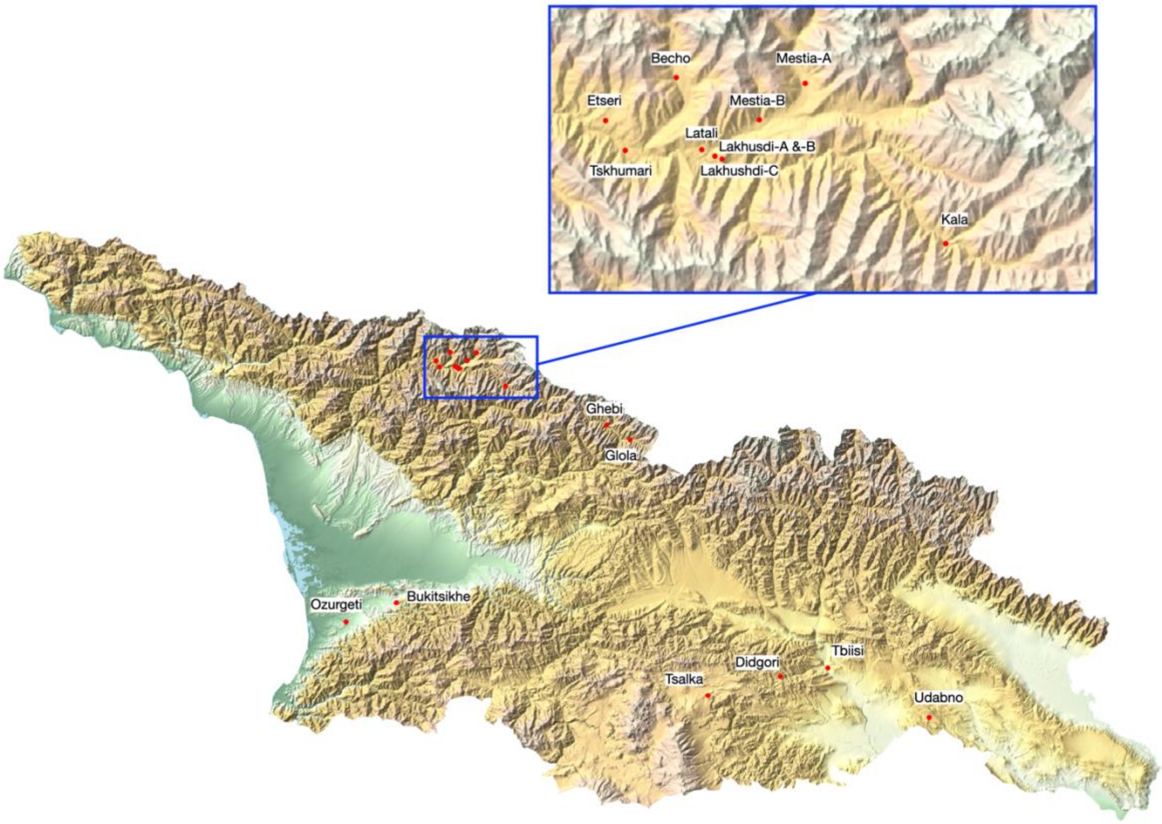


Figure 9. Recording locations of the different ensembles used in the present study. The villages of Ghebi and Glola are now located in a part of Racha which in former times was part of Svaneti. Ozurgeti and Bukitsikhe are located in Guria. All of the remaining locations, except for Georgia's capital Tbilisi correspond to villages in Svaneti or resettlements of Svan villages (Tsalka, Didgori, and Udabno).

As can be seen in Fig. 9, except for the villages of Tsalka, Didgori, and Udabno, which are eco-migrant Upper Svan communities populated a few decades ago, all of the other recordings of Svan music were obtained at locations in Upper Svaneti.

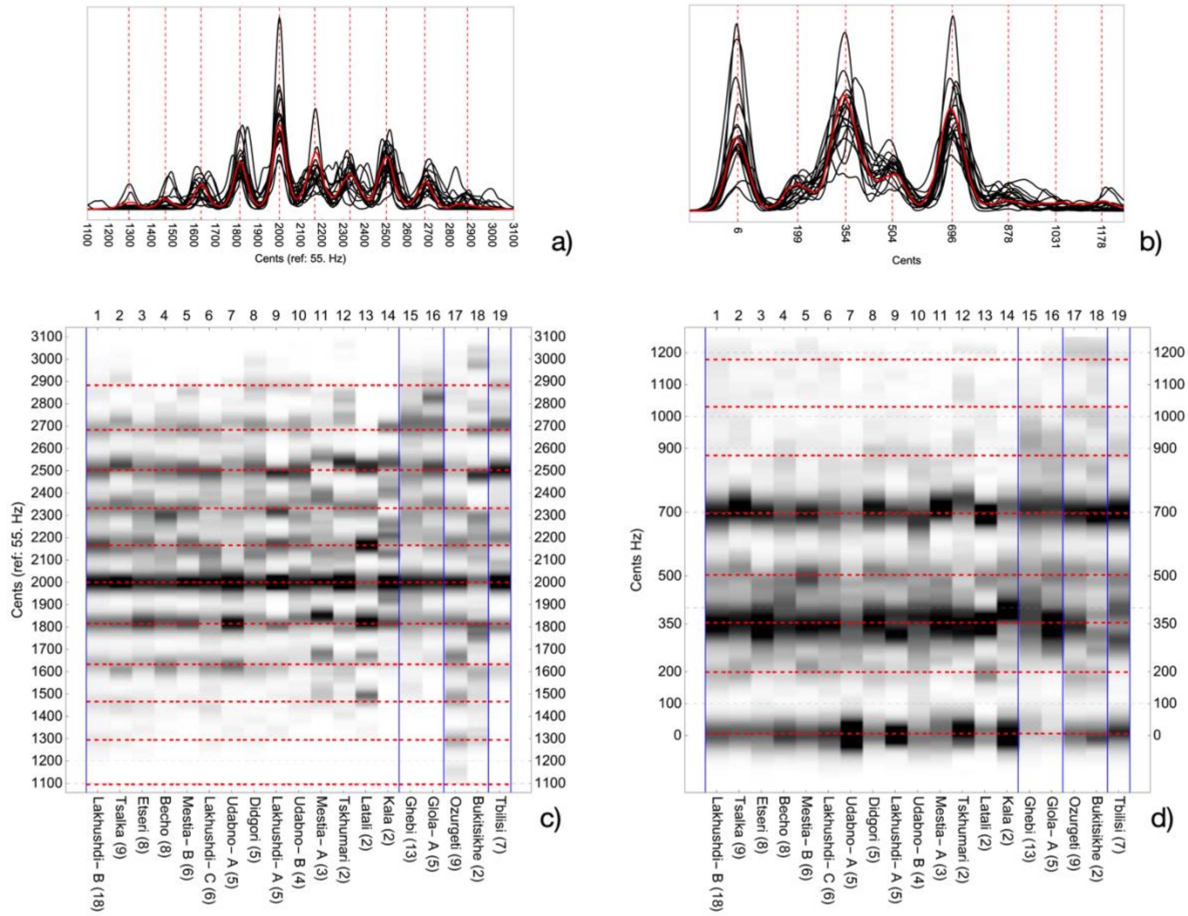


Figure 10. Average pitch and harmonic interval distributions for all ensembles (for the recording locations see Fig. 9). The red solid curves in (a) and (b) correspond to the average distributions calculated in the same way as in Fig. 6a and Fig. 7a. The vertical dashed lines in all four panels correspond to the pitch and interval group mean values for Lakhushdi-B and are shown here only for comparison.

Fig. 10 shows the comparison of all average ensemble-related pitch and harmonic interval distributions for the complete dataset. Since the recorded songs differ considerably for the individual ensembles, the corresponding pitch-group ranges (ambitus) shown in Fig. 10c are quite different. Nevertheless, the structures of the pitch distributions shown in Fig. 10a are quite similar in terms of spacing of the pitch groups. The similarities are even stronger in the case of the harmonic interval distributions, which are shown in the right panels of Fig. 10. On closer inspection and by repeatedly dividing the data sets, the final results of which are shown in Fig. 11, we found that harmonic interval distributions for all the Svan ensembles can be grouped in three slightly different groups, only differing in some details.

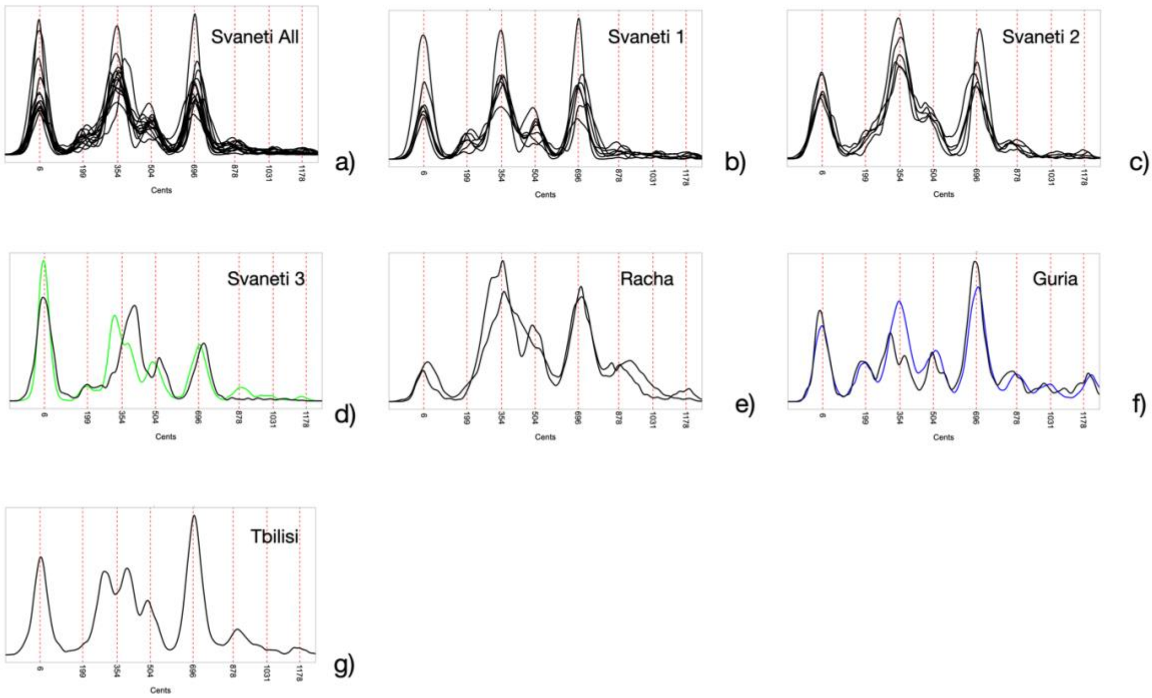


Figure 11. Average harmonic interval distributions for all ensembles and according to different groupings: (a) shows the superposition of all 14 harmonic interval distributions; (b) contains ensembles 1,2,5,7,8,12,13 and represents 43 recordings; (c) represent 29 recordings from the ensembles 3,4,6,10,11; (d) corresponds to the recordings from ensembles 9 (solid green line) and 14, representing a total of 7 recordings; (e) to (f) correspond to the ensembles from Racha, Guria, and Tbilisi, respectively, and are shown for comparison.

Subset Svaneti 1 in Fig. 11 b, for example, shows those ensembles where the seconds, thirds, and fourths are clearly separated. In contrast, there is less of a separation in the group Svaneti 2 (although the peaks are at the same locations). The separation between these interval groups disappears even more in one of the two Rachan ensembles (Ghebi). The two ensembles in Group Svaneti 3 are interesting because, in contrast to all the other Svan ensembles for which the harmonic thirds are distributed between 300 and 400 cents with a peak in the middle, one can identify a tendency to use minor and major thirds, in particular for ensemble 9 (Lakhushdi-A, green solid line). In this particular case, it may be explainable by the fact that they are also familiar with so-called city repertoire which often includes instruments in western tuning.

Overall, in terms of the harmonic tunings systems used, the group Lakhushdi-B seems to be a good representation of all the Svan ensembles in the 2016 GVM dataset.

An interesting anecdotal observation is that the peaks of the harmonic interval distribution used by the ensemble from Ozurgeti are essentially the same as those used by the Svan group Lakhushdi-B, the group with the largest number of recordings in our dataset. Finally, it is worth pointing out that the group of singers from Tbilisi seems to use a 12-TET (12-tone-equal-temperament) tuning system since all the peaks of the harmonic interval distribution appear at integer multiples of 100 cents.

Pitch Drifts

Here we briefly report on the results of the observations of the gradual pitch drifts, for which we had to compensate in our analysis. The pitch drift compensation curves employed are displayed in Fig. 12. It needs to be emphasized that ensembles used gradual pitch drifts inconsistently. i.e., only for certain songs, but different ensembles did not always do so for the same songs[10]. For example, for the song "Jgragish", some of the ensembles performed a pitch drift and some not. Furthermore, neither of the Gurian ensembles ever used it.

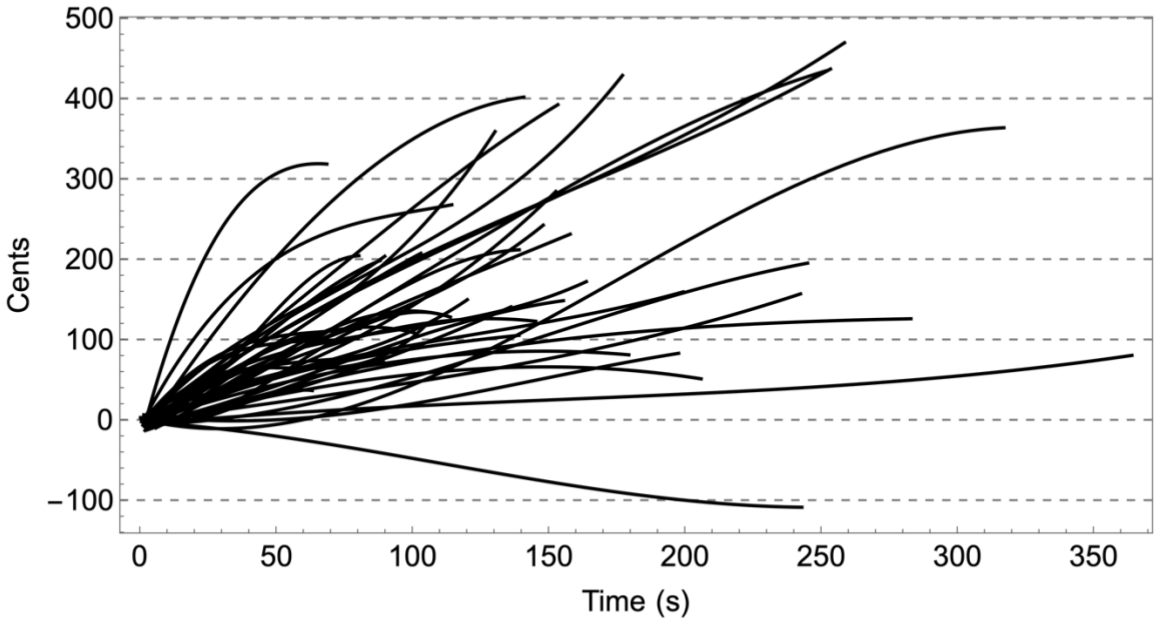


Figure 12. Pitch drift curves used for the compensation of the gradual pitch drifts observed for some of the recordings.

As shown in Fig. 12, there is only a single performance (from the group Udabno-A) where the pitch drift let to a decrease in pitch. All the other recordings of Svan singers showed an increase of up to approximately 400 cents in the course of a song, independent of its

[10] It would be interesting to know if the avoidance of pitch drifts could be related to singers having perfect pitch.

duration. It also seems, at first glance, that essentially two different slope groups were used, one with an average slope on the order of 30 cents per minute and a steeper one with an average slope on the order of 100 cents per minute.

Concluding Discussion

With the field expedition of 2016 and the collection of the 2016 GVM dataset, we tried to contribute to the documentation of the current performance practice of traditional vocal music in (Upper) Svaneti, the region which is believed to represent the oldest layer of Georgian multi-voice singing (Jordania, 2006). The present study aims to investigate what this dataset can contribute to the discussion of Georgian tuning systems. The approach is conceptually different from the analysis of historical audio recordings, in particular that of the Erkomaishvili dataset (Scherbaum et al., 2017; 2020; Rosenzweig et al. 2020), as it replaces the investigation of old recordings (which are usually of low fidelity) with the investigation of modern, high-quality recordings (but for which the link to the past is uncertain). Therefore, it is all the more interesting to compare the results of these two approaches. Because of the different recording setups and the different composition of the pitch inventories in the Erkomaishvili dataset and the 2016 GVM dataset in terms of the relative proportion of the contributing pitch groups, we compare only the mean values of the Gaussian mixture models for the melodic and harmonic tuning systems and ignore the associated standard deviations. The results of this comparison are listed in Table 1.

Table 1. Comparison of the synoptic scale models for the Erkomaishvili dataset from Fig. 23 (the one labelled all voices) of Scherbaum et al. (2020) with the average tuning systems obtained for all Svan ensembles. Scale degree 1 for the 2016 GVM melodic tuning system was calculated relative to the finalis pitch (2000 cents).

Scale degrees	AE melodic	2016 GVM melodic	AE harmonic	2016 GVM harmonic
8	1231	NA	1217	1182
7	1052	NA	1043	1018
6	886	868	874	868
5	705	693	707	703
4	509	509	515	495
3	342	332	355	349
2	176	163	191	205
1	0	0	0	6

Compared to a 12-TET (12-tone-equal-temperament) system, the Erkomaishvili and the Svan tuning systems are surprisingly similar, particularly for the smaller scale degrees.[11] Both systems show a strong presence of pure fourths and fifths, and 'neutral' thirds and sixths[12]. In addition, the sizes of the melodic and the harmonic seconds in both tuning systems differ with the size of the harmonic second (close to 200 cents) being systematically larger than the melodic one. This can be attributed, at least partially, to the harmonic second being a necessary side product of the popular 1-4-5 chord in traditional Georgian music.[13]

Finally, we want to emphasize that both our analysis and associated interpretation of the results are purely descriptive and have no normative aspects. None of the measures employed should be seen as a measure of any sort of 'quality', nor do we consider such a measure appropriate in the context of what we are trying to do. Our aim is simply to describe certain quantifiable aspects of current performance practices of traditional Georgian vocal music (with a focus on Svaneti and as they are reflected in the 2016 GVM dataset) in an as unbiased way as possible.

This notwithstanding, the current results raise serious questions regarding the practice of transcribing this music into a western notation system, in which neither the 'neutral' intervals nor the gradual pitch drifts can be appropriately represented. These observations underline the need for alternative unbiased representation forms of non-western music.

In the present work, we have deliberately addressed only those aspects of tonal organization on which conclusions can be drawn on the basis of the records available to us. This does not mean that we consider other aspects, such as the problem of "augmented

---

[11] Because of the smaller ambitus in Svan music, the higher degrees are less well represented in the GVM 2016 corpus.

[12] Essentially the same conclusion was reached by Mzhavanadze (2018) based, on the one hand, on the results of manual transcriptions of the Svan repertoire and, on the other hand, through a thorough comparative study of existing (published and archival manuscripts) notated transcriptions of Svan songs. It turned out that the same songs were transcribed differently by different scholars/musicians, which manifested itself in the different use of accidentals. These accidentals mainly come to the points with vertical harmonic thirds (while there is no divergence regarding harmonic fourths and fifths), which some seem to have interpreted as minor and others as major (to comply with the European notation system) thirds.

[13] This may not be the only explanation as we observed a performance of 'Tsmindao Ghmerto' by the ensemble Lakhushdi-B in which the top voice singer maintained a harmonic major second with the middle voice singer despite the fact that the fifth to the bass voice singer became considerably flat.

octaves" described by Georgian authors such as Karbelashvili or Gogotishvili (for a discussion see Jordania, 2006), unimportant. However, a quantitative, evidence-based study of this phenomenon, as well as a systematic comparison of the tuning systems of different regions, requires a much larger data set (including a sufficient number of octaves) and must be reserved for future work.

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 and the Fraunhofer Institut für Integrierte Schaltungen IIS. We thank David Gillman and Joseph Jordania for their constructive comments on an earlier version of the manuscript.

REFERENCES

Ambrazevičius, Rytis, Budrys, Robertas, & Viśnievska, Irena (2015). *Scales in Lithuanian traditional music: acoustics, cognition, and contexts*. Kaunas: Kaunas University of Technology.

Gómez, Emilia, Herrera, Perfecto, & Gómez-Martin, Francisco (2013). "Computational Ethnomusicology: perspectives and challenges" *Journal of New Music Research*, 42(2): 111–112.

Heller, Eric J. (2012). "Pitch Perception". In *Why You Hear What You Hear: An Experiential Approach to Sound, Music, and Psychoacoustics,* pp. 437 - 479. Princeton: Princeton University Press.

Jordania, Joseph (2006). *Who asked the first question? The origins of human choral singing, intelligence, language and speech*. Tbilisi: Logos.

Mauch, Matthias, Cannam, Chris, Bittner, Rachel, Fazekas, Giörgy, Salamon, Justin, Dai, Jiajie, Bello, Juan, Dixon, Simon (2015). Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency. *Proceedings of the First International Conference on Technologies for Music Notation and Representation* (TENOR), Paris, France.

Retrieved from http://www.tenor-conference.org/proceedings/2015/04-Mauch-Tony.pdf Last accessed 2022/12/09

Mauch, Matthias, & Dixon, Simon (2014). PYIN: A fundamental frequency estimator using probabilistic threshold distributions. Proceedings of the *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Florence/Italy. (pp. 659-663). Retrieved from https://ieeexplore.ieee.org/document/6853678 Last accessed 2022/12/09

Müller, Meinard, Rosenzweig, Sebastian, Driedger, Jonathan, & Scherbaum, Frank (2017). Interactive Fundamental Frequency Estimation with Applications to Ethnomusicological Research. Paper presented at *Conference on Semantic Audio, Erlangen, Germany*, *22 – 24 June 2017.*

Mzhavanadze, Nana (2018). "*Svanuri sak'ult'o rit'ualis musik'ologiur-antrop'ologiuri asp'ekt'ebi" (Musicological and anthropological aspects of Svan sacred ritual)*. PhD Dissertation, Ilia State University, Tbilisi: Georgia

Mzhavanadze, Nana, & Scherbaum, Frank (2020). "Svan Funeral Dirges (Zär): Musicological analysis" *Musicologist*, *4*(2): 168–197.

Rosenzweig, Sebastian, Scherbaum, Frank, & Meinard Müller. (2022). "Computer-Assisted Analysis of Field Recordings: A Case Study of Georgian Funeral Songs". *ACM Journal on Computing and Cultural Heritage* (JOCCH). Retrieved from https://dl.acm.org/doi/abs/10.1145/3551645

Rosenzweig, Sebastian, Scherbaum, Frank, & Müller, Meinard (2019). Detecting Stable Regions in Frequency Trajectories for Tonal Analysis of Traditional Georgian Vocal Music. Proceedings of the *20th International Society for Music Information Retrieval Conference.(Paper presented at 20th International Society for Music Information Retrieval Conference on November 4-8, 2019, Delft, The Netherlands) Flexer, Arthur; Peeters, Geoffroy; Urbano, Julian; Volk, Anjo (Eds.),* (pp. 352–359)*.* Retrieved from https://archives.ismir.net/ismir2019/paper/000041.pdf

Rosenzweig, Sebastian, Scherbaum, Frank, Shugliashvili, David, Arifi-Müller, Vlora, & Müller, Meinard (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):31–41.

Scherbaum, Frank, Müller, Meinard, & Rosenzweig, Sebastian (2017). Analysis of the Tbilisi State Conservatory Recordings of Artem Erkomaishvili in 1966. In *7th International Workshop on Folk Music Analysis, June 14-16 2017, Malaga, Spain* (pp. 29–36). Retrieved from https://doi.org/doi.org/10.25932/publishup-47614

Scherbaum, Frank, & Mzhavanadze, Nana (2020). Svan Funeral Dirges (Zär): Musical Acoustical Analysis of a New Collection of Field Recordings. *Musicologist*, *4*(2), 138–167.

Scherbaum, Frank, Mzhavanadze, Nana, Arom, Simha, Rosenzweig, Sebastian, & Müller, Meinard (2020). "Tonal Organization of the Erkomaishvili Dataset: Pitches, Scales, Melodies and Harmonies" *Computational Analysis of Traditional Georgian Vocal Music*. Ed. Frank Scherbaum: pp. 64. Potsdam/Germany: Universitätsverlag Potsdam.

Scherbaum, Frank, Mzhavanadze, Nana, & Dadunashvili, Elguja (2018). „A Web-Based, Long-Term Archive of Audio, Video, and Larynx-Microphone Field Recordings of Traditional Georgian Singing, Praying and Lamenting with Special Emphasis on Svaneti" Paper presented at *9. International Symposium on Traditional Polyphony,* Tbilisi, Georgia, Organization date.

Scherbaum, Frank, Mzhavanadze, Nana, Rosenzweig, Sebastian, & Müller, Meinard (2019). "Multi-media recordings of traditional Georgian vocal music for computational analysis". Paper presented at the *9th International Workshop on Folk Music Analysis,* Birmingham, England. *2 -4 July, 2019*.

Tzanetakis, George (2014). "Computational ethnomusicology: a music information retrieval perspective" Proceedings of the International Computer Music Conference ICMC|SMC|2014, 14-20 September 2014, Athens/Greece, pp. 69–74.

Tzanetakis, George, Kapur, Ajay, Schloss, Walter Andrew Schloss, & Wright, Mathew (2007). "Computational Ethnomusicology" *Journal of Interdisciplinary Music Studies*, 1(2): 1–24.

Wolfram, Stephen. (2016). How to teach computational thinking. Blogpost "Stephen Wolfram Writings", Retrieved from https://writings.stephenwolfram.com/2016/09/how-to-teach-computational-thinking/ (Last accessed 2022/09/15).