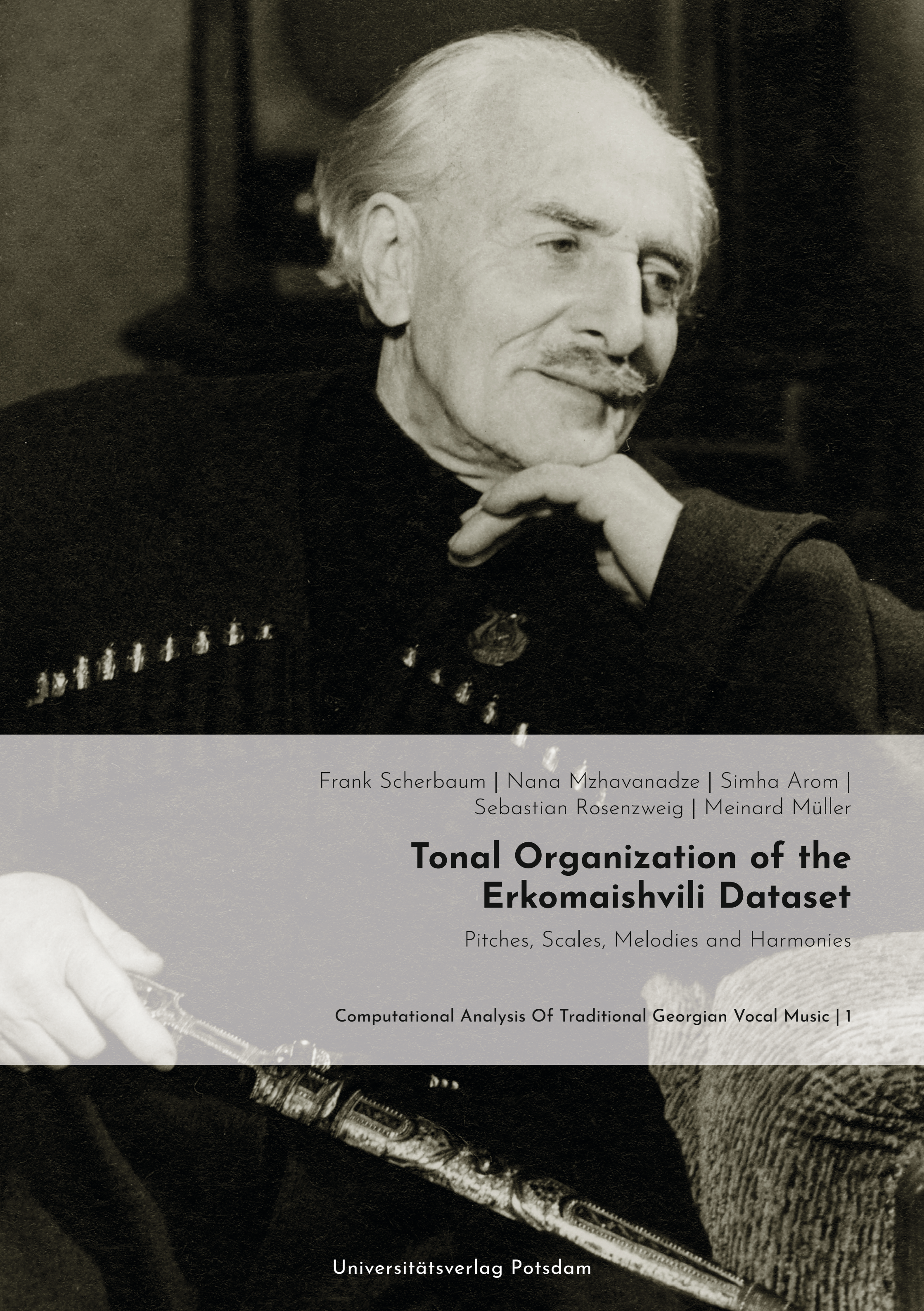

Frank Scherbaum | Nana Mzhavanadze | Simha Arom |
Sebastian Rosenzweig | Meinard Müller

Tonal Organization of the Erkomaishvili Dataset

Pitches, Scales, Melodies and Harmonies

Computational Analysis Of Traditional Georgian Vocal Music | 1

Universitätsverlag Potsdam

Computational Analysis Of Traditional Georgian Vocal Music

Computational Analysis Of Traditional Georgian Vocal Music | 1

Frank Scherbaum, Nana Mzhavanadze, Simha Arom,
Sebastian Rosenzweig, Meinard Müller

Tonal Organization of the Erkomaishvili Dataset: Pitches, Scales, Melodies and Harmonies

Universitätsverlag Potsdam

Frank Scherbaum University of Potsdam, Potsdam, Germany
Nana Mzhavanadze University of Potsdam, Potsdam, Germany
Simha Arom CNRS Paris, France
Sebastian Rosenzweig International Audio Laboratories Erlangen, Germany
Meinard Müller International Audio Laboratories Erlangen, Germany

Bibliografische Information der Deutschen Nationalbibliothek
Die Deutsche Nationalbibliothek verzeichnet diese Publikation in der Deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über http://dnb.dnb.de/ abrufbar.

Universitätsverlag Potsdam 2020
http://verlag.ub.uni-potsdam.de/

Am Neuen Palais 10, 14469 Potsdam
Tel.: +49 (0)331 977 2533 / Fax: -2292
E-Mail: verlag@uni-potsdam.de

Die Schriftenreihe **Computational Analysis Of Traditional Georgian Vocal Music** wird herausgegeben von Prof. Dr. Frank Scherbaum

ISSN (online) N. N.

Umschlagfoto: Artem Erkomaishvili: aus der privaten Sammlung von Anzor Erkomaishvili
Satz: text plus form, Dresden

Zugleich online veröffentlicht auf dem Publikationsserver
der Universität Potsdam:
https://nbn-resolving.org/urn:nbn:de:kobv:517-opus4-476141
https://doi.org/10.25932/publishup-47614

Contents

In this study we examine the tonal organization of a series of recordings of liturgical chants, sung in 1966 by the Georgian master singer Artem Erkomaishvili. This dataset is the oldest corpus of Georgian chants from which the time synchronous F0-trajectories for all three voices have been reliably determined (Müller et al. 2017). It is therefore of outstanding importance for the understanding of the tuning principles of traditional Georgian vocal music.

The aim of the present study is to use various computational methods to analyze what these recordings can contribute to the ongoing scientific dispute about traditional Georgian tuning systems. The starting point for the present analysis is the re-release of the original audio data together with estimated fundamental frequency (F0) trajectories for each of the three voices, beat annotations, and digital scores (Rosenzweig et al. 2020). We present synoptic models for the pitch and the harmonic interval distributions, which are the first of such models for which the complete Erkomaishvili dataset was used. We show that these distributions can be very compactly expressed as Gaussian mixture models, anchored on discrete sets of pitch or interval values for the pitch and interval distributions, respectively. As part of our study we demonstrate that these pitch values, which we refer to as *scale pitches,* and which are determined as the mean values of the Gaussian mixture elements, define the scale degrees of the melodic sound scales which build the skeleton of Artem Erkomaishvili's intonation. The observation of consistent pitch bending of notes in melodic phrases, which appear in identical form in a group of chants, as well as the observation of harmonically driven intonation adjustments, which are clearly documented for all pure harmonic intervals, demonstrate that Artem Erkomaishvili intentionally deviates from the scale pitch skeleton quite freely. As a central result of our study, we prove that this melodic freedom is always constrained by the attracting influence of the scale pitches. Deviations of the F0-values of individual note events from the scale pitches at one instance of time are compensated for in the subsequent melodic steps. This suggests a *deviation-compensation* mechanism at the core of Artem Erkomaishvili's melody generation, which clearly honors the scales but still allows for a large degree of melodic flexibility. This model, which summarizes all partial aspects of our analysis, is consistent with the melodic scale models derived from the observed pitch distributions, as well as with the melodic and harmonic interval distributions. In addition to the tangible results of our work, we believe that our work has general implications for the determination of tuning models from audio data, in particular for non-tempered music.

1 Introduction

The rich musical heritage of the country of Georgia has long attracted the attention of musicians, music lovers, and musicologists. After the recognition of its traditional polyphonic singing as an intangible cultural heritage by UNESCO in 2001 and the establishment of the bi-annual International Symposia on Traditional Polyphony in 2002 at the State Conservatory in Tbilisi, the research on Georgian traditional music has taken on a new momentum. One of the most controversial issues of discussion in this context continues to be the (authentic) structure of the tonal organization of traditional Georgian vocal music. Some scholars, e.g. (Erkvanidze 2016), have argued that the key to understanding this issue lies in the analysis of old audio recordings from professional singers of the last century. Although the first recordings of traditional Georgian singing with phonographs date back as far as the beginning of the last century, the poor audio quality, caused by an extremely high noise level and strong signal distortions due to the recording process, makes analysis with automated methods practically impossible. To our knowledge, the oldest set of recordings of reasonable quality and of sufficient size for drawing general conclusions are the recordings by Artem Erkomaishvili from the year 1966 (recorded at the Tbilisi State Conservatory). This dataset, which is the object of the present study, is believed to be an essential source for any theory on tuning principles of traditional Georgian vocal music (see Graham 2015). Artem Erkomaishvili (1887–1967) was one of the last professional master chanters in Georgia and a giant of traditional Georgian vocal music of the 20th century. Until today, his family name stands for a long list of famous singers and choir leaders from the region of Guria in Western Georgia, a list which reaches back far into the 19th century (Erkomaishvili 2017). With the Tbilisi State Conservatory recordings of 1966, of which a remastered selection has recently been released under the name Pearls of Georgian Chant (Jgharkava 2016), Artem Erkomaishvili has left a legacy which is invaluable in a number of ways. Since chanting was prohibited during the Soviet period, without these recordings the tradition of the Shemokmedi chanting school might actually not have survived the last century (Shugliashvili 2014).

The objective of the present study is to investigate the tonal organization of the Erkomaishvili dataset from several different perspectives. What we consider first in the context of our analysis are the statistical distributions of the F0-values (i.e. the values determined by the pitch tracking algorithms used by us) in those parts of the F0-trajectories that are assumed to be perceived as note objects with well-defined pitches. We will refer to them as pitch[1] distributions and all F0-values within note events as pitches or pitch samples. Hence, the term pitch in the present paper should be understood as an acoustic quantity, which at best can be interpreted as a proxy for the psychoacoustic quantity, which can not be measured directly. Modeling the pitch distributions by so-called Gaussian mixture distributions, which will formally be introduced in Section 4.2, will lead to a very efficient numerical representation as well as to melodic sound scale models. Subsequently, we investigate the tonal organization through what we refer to as *melodic step size distributions,* which are the distributions of F0-differences

1 In the field of psychoacoustics, the term ‘pitch’ is usually 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. Those algorithms will return F0-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.

between consecutive note events in a melody. The third perspective is via the distribution of *harmonic intervals* in those parts of F0-trajectories from different voices which belong to note events and for which concomitant F0-values are available. Finally, we analyze the microtonal structure of the Erkomaishvili dataset (at sub-semitone resolution) and investigate the phenomenon of dynamic intonation adjustments for individual note events and different interval types. The general goal of our study is to develop an evidence-based model for Artem Erkomaishvili's melody generation process which incorporates all these aspects of tonal organization (pitches, scales, melodies, and harmonies) in a consistent manner. The paper is structured as follows. After this introduction, we review in Section 2 the publicly available literature on tonal organization of traditional Georgian music, thus introducing the context of the current work. Section 3 presents the main aspects of the Erkomaishvili dataset as far as they are relevant for the present study. Section 4 explains our overall analysis strategy and introduces the methodologies used in the subsequent analysis. We then turn to the actual processing of the recordings. We first analyze the top voice recordings (which are not affected by the physical presence of other voices) in Section 5 and then, in Section 6, their combination with the middle and bass voices (which were sung against pre-recorded voices). Section 6 closes with the discussion of the harmonic interval distributions. The microtonal structure of the dataset and the phenomenon of harmonic intonation adjustments are discussed in Section 7, followed by a discussion and conclusions in Section 8.

2 Discourse on the Tonal Organization of Traditional Georgian Music

In this section, we present an overview of previous related studies on the tonal organization of traditional Georgian music. For some of these studies, the lack of documentation makes it difficult to judge their informative value according to modern scientific standards. However, in order to achieve as complete a picture as possible, we include all accessible publications in the following review. A collection of works by Georgian authors has been translated and published for an international audience by Tsurtsumia and Jordania (2010). These studies go well beyond tuning and scale issues, also covering geographical, historical and cultural aspects, which are beyond the focus of our present study. This notwithstanding, these publications to some degree also discuss aspects related to the tonal organization of the music. Gogotishvili (2004a), for example, describes the structure of the typical scale systems in traditional Georgian polyphony as non-octave scales being either fourth-based, fifth-based or based on a mixture of both. One of the key features of the fifth-based system is the augmented octave and augmented fourth (tritone). In contrast, fourth-based scales would be tritone free. Other authors such as Aslanishvili (2010), Araqishvili (2010), Chkhikvadze (2010), Jordania (2010) emphasize the strong role of the (parallel) fifth and the 1-4-5 chord as being a characteristic feature of Georgian traditional music. Another aspect, which by some musicologists has been considered characteristic for Georgian singing, is the influence of harmonic constraints on the fine-tuning of the voices (see Nadel 1933; Chokhonelidze 2010).

While it can be considered consensus amongst researchers that historically, traditional Georgian vocal music was not tuned to the 12-tone equal-temperament scale, the particular nature of the Georgian sound scale(s) is still an ongoing topic of intense discussion. Erkvanidze (2002, 2016), for example, claims that the historical Georgian sound scale is related to the ancient Greek modal system. Based on this assumption and acoustic measurements, which in his first paper he describes as having being done “by ear” (Erkvanidze 2002), he proposed tuning models based on two different combinations of two tetrachords in which the interval sizes for the second can take on three different values, namely 172, 154, and 204 cents. In contrast, based on the spectral analysis of selected melodic fragments from historical audio recordings, Tsereteli and Veshapidze (2014) propose a sound scale model in which all the melodic intervals are assumed to be of equal size which is assumed to be 1200/7 = 171.4 cents. They emphasize, however, that this value should not be taken as a rigid quantity. Traditional Georgian singers, similar to the use of *Blue Notes* in Blues or Jazz, would commonly deviate from this value by some amount as part of their personal interpretation. This pitch bending, as it is commonly called for other musical styles, is referred to as using *shinpardis*[2] (see Tsereteli and Veshapidze 2014; Erkvanidze 2016).

Another equal-interval-size model, though with an interval size of 700/4 = 175 cents, was suggested by Gelzer (2002). His model was derived by iteratively minimizing the mismatch of computer-generated synthetic sound scales with the pitches he heard in recordings. Conceptually, his model, as well as the model by Gogotishvili (2004b), are built on fifths, rather than on octaves, which leads to a scale in which notes separated by one or more octaves are musically not equivalent (missing octave equivalence).

In contrast to the equal-interval-size models, Kawai et al. (2010) analyze a single Megrelian song from a teaching CD and vaguely describe the sound structure of this example as being based on "unequal but variable interval sizes". Westman (2002) takes a completely different perspective on the problem of tonality. He questions the usefulness of the scale concept to describe tonality (in particular to identify a tonic) in Georgian music altogether and suggests a tonality model in which interval, pitch, and timbre interact. He suggests that microtonal variations could be either the product of a scale, or an interval for melodic purposes, or the product of pitch expression. However, he acknowledges that how these parameters might interact needs further study.

One of the major challenges in the context of the Georgian sound scale discourse, at least as it is reflected in the internationally accessible literature discussed above, is the fact that previously only very small datasets have been considered as observational evidence for the

2 The term *shinpardi* originally stems from Ione Batonishvili's "Kalmasoba" (http://buki.ge/library-5527.html, last accessed on: August 24, 2020) where it says "In every country … Greece, Europe, Arabia … (and in our country, too) the chants are divided into eight voices (echoi). You can't divide them anymore, because the frets on the instruments are also eight, which when playing make different sounds. And the K'ank'ledi (კანკლედი), which is called Shinpardi (შინფარდი), makes some slight changes within the (frets). And it is used in church singing in the same way as in instruments. And it also comes from eight voices (echoi)." Translated by Nana Mzhavanadze.

suggested models. Furthermore, the technical details of the earlier studies are only sparsely documented. It often remains unclear if the reasons for the differences between these models are due to the usage of different methods and/or different datasets or by incompatibility of (undocumented) underlying assumptions. With the exception of the study of Tsereteli and Veshapidze (2014), for which technical details of the analysis are provided in the video of a conference presentation (Tsereteli and Veshapidze 2015), none of the earlier studies seemed to have put a lot of emphasis on trying to make their study transparent or reproducible.

3 The Erkomaishvili Dataset

In 1966, one year before his death, the aging Artem Erkomaishvili was recorded by ethnomusicologist Kakhi Rosebashvili performing three-voice chants at the Tbilisi State Conservatory. Due to the lack of fellow singers, he had to sing all three voices by himself. This was made possible by sequentially recording the three voices using an overdubbing technique, leading to three subsequently recorded temporal segments, as illustrated in Figure 1 (a). The top voice, which in Georgian chant is also the leading voice, was recorded first (as solo voice), then played back to him while he was singing the middle voice. Finally, he sang the bass voice against the playback of the superposition of the top and middle voice.

In general, the determination of pitches from polyphonic audio recordings is currently still unsolved, in particular from old recordings with low dynamic range and bandwidth. In the case of the Erkomaishvili recordings, however, the sequential recording technique allowed the reconstruction of the F0-trajectories for all three voices as well as some preliminary analyses (Müller et al. 2017; Scherbaum, Müller, and Rosenzweig 2017a, 2017b; Rosenzweig, Scherbaum, and Müller 2019).

As of today, the Erkomaishvili dataset is the oldest set of recordings of Georgian chants from which the time synchronous F0-trajectories (sequences of F0-values sampled every 5.8 msec (Müller et al. 2017)) for all three voices have been reliably determined. Although original-

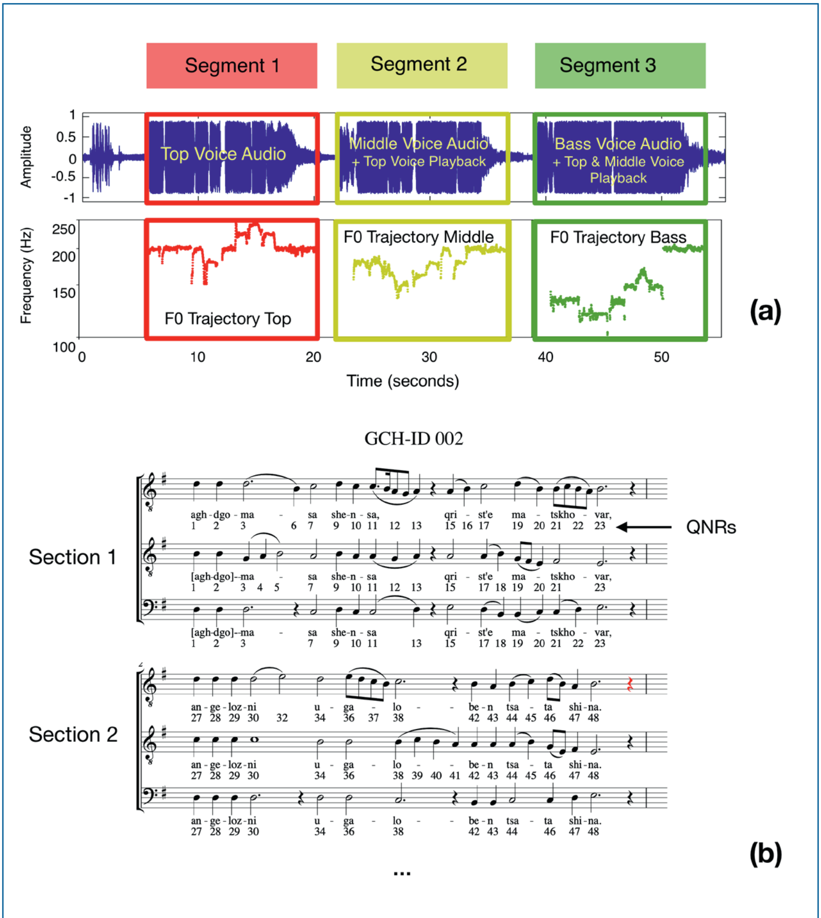

Figure 1: The Erkomaishvili dataset. (a) Three-segment recordings with F0-annotations. (b) Digital transcription of the song with GCH-ID 002. GCH-ID 002 refers to the chant ID system of (Rosenzweig et al. 2020) which is also used throughout this paper. The chants are temporally subdivided into different sections (*mukhli*). QNR indicates the quarter note references. The time axis in Figure 1(a) starts from the beginning of the top voice segment and ends at the end of the bass voice segment.

ly caused by the absence of fellow singers, the sequential recording of the individual voices of Artem Erkomaishvili against the playback of his own voice(s) turns out to be a significant advantage for the analysis. First, it enables the analysis of melodic aspects for the top voice using the monophonic audio signals from the first recording segment as well as from the corresponding F0-trajectory.

Furthermore, the second segment with two superimposed voices and the third segment with three superimposed voices allow for studying harmonic intervals, and harmonic aspects such as the influence of harmonic interval perception, from the F0-trajectories.

The dataset was recently re-released in a newly organized and manually annotated form, together with an interactive web-based interface. Using score-following audio players that make use of the annotated data, the interface provides direct and convenient access to the corpus (Rosenzweig et al. 2020). The recorded chants were transcribed in a manual process by (Shugliashvili 2014). Through the websites provided by Rosenzweig et al. (2020), the publicly available audio material along with all fundamental frequency annotations, recording structure annotations, XML versions of the transcriptions, and note onset annotations were made available.[3] As a result, the source material for the analysis discussed in the present paper is publicly accessible.

In the present study, we make particular use of the manually generated onset annotations based on the quarter note reference (QNR) system introduced by Rosenzweig et al. (2020), see Figure 1 (b). The QNR system aligns the audio recordings with the digitized score information. This makes it possible to computationally compare the transcribed musical scores with acoustical properties of Erkomaishvili's recorded performances. This, in turn, enables the quantitative investigation of the tonal organization of the whole corpus in an unprecedented form, e.g. the investigation of harmonic tuning adjustments for individual notes (cf. Section 7.2).

3 https://www.audiolabs-erlangen.de/resources/MIR/2019-GeorgianMusic-Erkomaishvili.

For the analysis in the following sections, some peculiarities of the corpus need to be mentioned. First, the temporal structure of chants is non-metrical in the sense that the concept of measures does not apply. Instead, chants are usually subdivided in temporal units called *mukhli* in Georgian. We will refer to them by their Georgian name or as sections (to distinguish them from the segments of the recording process). An example is shown in Figure 1 (b). Second, some groups of chants are similar in their melodic and harmonic structures. For example, the group of chants with GCH-IDs 008, 009, 010, and 011, which we will refer to as CG 008-011, contains identical melodic phrases consisting of up to 10 quarter notes. In order to understand which factors influence Artem Erkomaishvili's voice production, we will study this group of chants in great detail. To this end, we will examine in Section 5.1.2 to what extent the individual realizations of these phrases differ from each other and to what extent Artem Erkomaishvili seems to intentionally deviate from an underlying fixed scale (pitch bending). In addition, the analysis of pitch fluctuations across section boundaries will be examined to quantitatively assess the amount of unintended pitch fluctuations in his vocal production (cf. Section 4.2).

4 Analysis Strategy

Before we turn to the actual analysis of the recordings in Section 5, we first explain our overall analysis strategy and discuss some methodological aspects. In the context of music perception, it has been suggested to distinguish between the sequential (horizontal, melodic) and the concomitant (vertical, harmonic) structure (Nikolsky 2015). For instruments with fixed pitches, the sequential intervals in a melody (melodic intervals) and the concomitant intervals in a chord (harmonic intervals) are elements of the same interval set. For a-cappella vocal music, however, this is not necessarily the case and pitch distributions as well as melodic and harmonic interval distributions can become rather different. Therefore, in the subsequent analysis, we treat them as separate entities.

In Section 4.1, we describe how we approach the determination of pitch distributions from both audio tracks and F0-trajectories, and briefly touch upon their interpretation against the background of known phenomena of pitch perception. We then discuss in Section 4.2 the determination and interpretation of melodic sound scales from pitch distributions and how one can test if these have a normative or only a descriptive value for Artem Erkomaishvili's voice production. Finally, in Section 4.3, we address the analysis of harmonic intervals and introduce a novel approach to estimate the effect of harmonic intonation adjustments for individual notes.

4.1 Determination of Pitch Distributions

As an initial step of the melodic analysis, we determine what we refer to as *pitch distributions*. These were defined in Section 1 as the distributions of all F0-values from those parts of the F0-trajectories which can be associated with note events with well-defined pitches. Depending on the available material, the computation of pitch distributions can be done in different ways. For the top voices, for which we have the monophonic audio recordings, see Figure 1 (a), we can make use of highly sophisticated and well-tested software tools such as Tony (Mauch et al. 2015), which allows to visually and acoustically control precisely which parts of an audio signal are used for the determination of note events. This process is very time consuming and will therefore only be used for the analysis of the chant group CG 008-011 in Section 5.1, where we want to achieve the highest possible control over which parts of a signal are used for pitch determination. As still reliable, but timewise less demanding alternatives, there exist two approaches for note event detection in F0-trajectories that have recently been proposed (Rosenzweig, Scherbaum, and Müller 2019). These approaches aim at finding approximately stable horizontal parts in a F0-trajectory. The first algorithm uses morphological operations inspired by image processing (Vavra et al. 2004), while the second is based on suitably generated binary time-frequency masks. To avoid undesired distortions in subsequent analysis steps, both approaches keep the original F0-values unmodified, while only removing F0-values in unstable trajectory regions. In our study, we use the morphological approach with a filter range of 10 points (corresponding to 58 msec) and a dynamic threshold value which is calculated for each trajectory based on the assumption made after visual inspection of a number of chants that 70 percent of the F0-trajectories belong to note segments (see Rosenzweig, Scherbaum, and Müller 2019 for details).

4.1.1 Influence of Pitch Perception

After computing pitch distributions in which all pitch samples belong to note objects with perceptionally well-defined pitches, we still face a whole collection of challenges regarding the interpretation of these pitches in terms of an underlying tuning system or a sound scale. The situation is complicated because singers and listeners are exposed to a whole series of psychological phenomena that influence both the perception and production of a vocal sound. These include the phenomenon of categorical perception (CP), which leads to the observation that listeners (especially musicians) seem to have a higher sensitivity in distinguishing pitches at the border of different pitch categories, but sometimes a lower sensitivity within a category (perceptual magnet effect). For an excellent discussion of the most important aspects of these phenomena related to vocal music we refer to Ganguli and Rao (2019).

4.1.2 Determination and Interpretation of Melodic Sound Scale Models

In general, the pitches in the pitch distribution appear to be strongly clustered, with each of the pitch clusters, which we will also refer to as pitch groups or simply as degrees, showing a roughly symmetrical shape (cf. histogram in Figure 2(b)). This justifies modeling them as Gaussian Mixture Models (GMMs), which are simply weighted mixtures of individual Gaussian distributions $\mathcal{N}(\mu, \sigma^2)$ each of which is defined by a mean value μ and a standard deviation σ. For the case of K pitch groups, this results in a representation as $\sum_{k=1}^{K} w_k\, \mathcal{N}(\mu_k, \sigma_k^2)$.

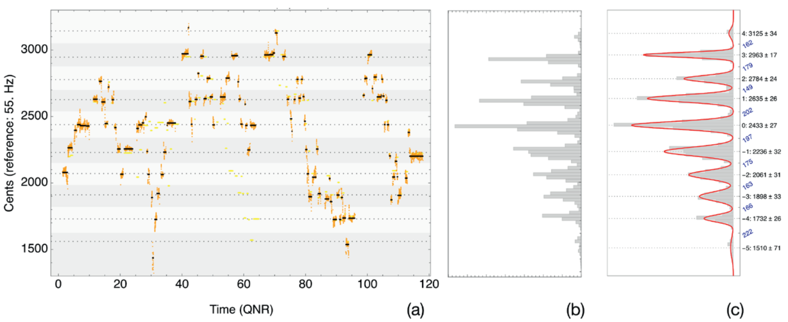

Figure 2: Determination of a pitch distribution and melodic sound scale for the top voice of chant GCH-ID 008. (a) Note objects (short black horizontal lines) and corresponding pitch track (yellow) were determined using the Tony software (Mauch et al. 2015). (b) Pitch histogram. (c) Corresponding Gaussian mixture distribution.

The mean values of the individual Gaussians (the μ_k) which correspond to the center values of the individual pitch groups, are assumed to define 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 reason for this variability can be a combination of the effects of categorical perception, the perceptual magnet effect, random fluctuations, but also of intended changes of individual pitches as a matter of the singer’s personal style of expression. How can these phenomena be distinguished? How can one tell from the recordings when a singer uses *shinpardis* (in terms of Tsereteli and Veshapidze (2014)) and when a singer is tired or reaches the precision limits of his voice control? And how can one distinguish the effects of pitch fluctuations from the effects of categorical perception and perceptual magnet effects? Fortunately, for this purpose, we can benefit from peculiarities of the Erkomaishvili dataset described in Section 3, at least to determine the order of magnitude of these effects.

Specifically, we can make use of the fact that the chants are structured in sections (*mukhli*) and second, of the existence of groups of similar chants which contain identical melodic phrases. The melodic progression in chants usually occurs in small melodic steps (usually second intervals) except at the boundaries of sections. Here, sometimes larger melodic intervals, such as fifths or even octaves occur. These melodic jumps are usually associated with small differences of the pitch distributions before and after the section boundary. We interpret these shifts, which are usually very small and which we assume to be unintended, as a random (noise) contribution to the pitch distribution of a chant and as a measure of the maximum achievable pitch resolution, see Section 5.1.1. Second, the effect of intended pitch bending will be studied through the detailed investigation of the chant group CG 008-011 in Section 5.1.2. The analysis of the identical phrases in the different chants allows us to determine to which degree pitches deviate from the underlying melodic scale model for individual notes and for all chants of the group.

4.2 Harmonic Analysis and Intonation Adjustments

In the second part of the analysis, we will concentrate on the harmonic aspects of the tonal organization. This adds another level of complexity to the interpretation of pitch shifts since harmonic interval perception is known to influence the fine-tuning of voices, which in turn can lead to unintended pitch shifts of a whole ensemble (see e.g. Howard 2007). In addition, intentional intonation adjustments have been suggested to be characteristic for traditional Georgian singing (Nadel 1933; Chokhonelidze 2010) and an expression of what sometimes is referred to as “vertical musical thinking”.

Our strategy to estimate the effect of harmonic intonation adjustments is as follows: From the onset times and the durations of the note events estimated from the F0-trajectories of the individual voices, we first calculate those time windows where we have concomitant pitch information on the combination of voices of interest. In other words, the onset times and durations of either harmonic intervals or, more general, chord objects. Subsequently, we determine the pitch trajectories for each voice of interest in these time windows. The average values of the differences of these short pitch trajectories (in cents) define the values of the harmonic intervals for the voice combination under consideration. By collecting all harmonic intervals present in a chant, we obtain what we refer to as harmonic interval inventory or harmonic interval distribution. Similar to the melodic pitch histograms, the harmonic interval histograms appear to be strongly grouped and can also be modelled by Gaussian Mixture Models (GMMs) (see Section 6.4), which will reflect the harmonic aspects of the tonal organization. Finally, the amount of correlation of the pitch fluctuations between the individual voices within the time windows of an interval or a chord object can be used to derive a quantitative measure for the degree of harmonic intonation adjustment. Simply speaking, intonation adjustment will force the pitch trajectories of two voices to become dependent on each other, which can be mathematically quantified. The theoretical derivation of this measure is given in Section 7.2.

Since beat annotations have become available for all the chants (Rosenzweig et al. 2020), harmonic intonation adjustments can now be studied on a note level time scale. This makes it possible to investigate this phenomenon for any phrase of interest in the whole corpus. In this article, however, we will limit ourselves to investigate if the average amount of harmonic adjustment differs for different intervals and leave the more detailed time-dependent analysis to future studies.

5 Top Voice Analysis

In this section, we turn to the actual pitch analysis of the recordings of the Erkomaishvili dataset and discuss the results of our data processing. For the first part, we restrict ourselves to the processing of the top voice recordings. The reason for the separate analysis of the top voices from the middle and bass voices, which will be discussed in Section 6.1, is that first, the top voice is the reference voice in liturgical chants and musically the most important one. Second, unaccompanied singing and singing in the presence of pre-recorded voices are different. Therefore, by analysing the top voice recordings separately, we want to separate melodic and harmonic aspects of the tonal organization as much as possible. Third, the top voice recordings are also special for a technical reason. In this case, we have both the audio and the F0-trajectories at our disposal. As already discussed in Section 4.1, this gives us more options to perform the pitch analysis for the top voice than for the middle and bass voices, where we only have the F0-trajectories to work with.

5.1 Analysis of Chant Group CG 008-011

In the following, we study the four chants belonging to the chant group CG 008-011 in detail. In particular, we will examine the relationship between the pitch distribution of the chant GCH-ID 008 and the corresponding distribution of melodic intervals. This will lead to a model which explains how the top voice melody is formed in relation to the sound scale derived from the pitch distribution. In addition, we will exploit the fact that the chant group CG 008-011 contains identical melodic phrases in order to investigate the effect of intentional pitch bending. Finally, we will make use of the musical subdivision of the chant into sections (*mukhli*) to examine the stability of the pitch distributions for different sections.

Figure 2(a) shows the pitch track of the top voice of chant GCH-ID 008 and Figure 2(b) the corresponding pitch histogram. In this case, the note events, shown by the horizontal black lines, are determined from the raw audio trace by using the Tony software (Mauch et al. 2015). The reason for using Tony for the analysis of chant group CG 008-011 is that we wanted to have the maximum degree of confidence that the pitch tracks have been reduced to those parts of the F0-trajectories which a listener would perceive as note objects with a perceptionally well-defined pitch. This was achieved by visually and acoustically checking and manually editing individual note events. This interactive post-processing capability of Tony is currently not available with the approach proposed by Rosenzweig, Scherbaum, and Müller (2019).

The pitch histogram in Figure 2(b) shows that the pitches are not evenly distributed along the vertical pitch axis but are strongly clustered and appear as individual pitch groups. A model which accommodates such a grouping is a Gaussian mixture model in which the pitch set, which in case of Figure 2 consists of 7675 values, is fit by a weighted mixture of K Gaussian distributions, see Figure 2(c). This is done by using a modified version of the of the expectation-maximization algorithm by Fox (2011).

The mean values μ_k and standard deviations σ_k (in cents) for each of the Gaussians are displayed to the right of each pitch group as $\mu_k \pm \sigma_k$, following the so-called pitch group indices PGI, which represent the distance to the most salient pitch group (which is defined to have a *PGI* of 0). For example, in Figure 2, we have $K = 10$ pitch groups, with the most salient one (the pitch group with the largest area under its Gaussian) being the 6th one ($k = 6$), if counted from lowest to highest pitch. Hence, PGI = 0 in this example corresponds to $\mu_6 = 2433$ and $\sigma_6 = 27$. The most salient pitch group corresponds to the pitch group which is most often heard in the complete chant. This in turn makes it perceptionally salient because its representation in short-term memory is enhanced in comparison to the other pitch groups present, as shown in a number of studies by (Deutsch 1970, 1972, 1975). Therefore, PGIs reference the pitch groups in a perceptionally meaningful way without having to make assumptions about functional relations (e.g. the relation to a tonic). Finally, in Figure 2(c), the tilted blue numbers between the pitch groups specify the differences (given in cents) between the center pitches of neighboring pitch groups. Loosely speaking, we refer to these differences as *expected* step sizes of melodic seconds. It is interesting to note that the expected step sizes between the most salient pitch group and its immediate neighbors are close to 200 cents (210 cents and 188 cents, respectively), while most of the others are smaller. In addition, the differences between the most salient pitch group and pitch groups −4 and 4 are both close to a pure fifth. As a consequence, pitch group −4, −1, 0 with its constituting intervals 499, and 709 cents builds a nearly pure 1-4-5 chord, as does the combination of pitch groups 0, 3, 4 with 507 and 704 cents, respectively.

Figure 3(a) shows the *actual* melodic seconds in the top voice of chant GCH-ID 008 as they were sung by Artem Erkomaishvili. The color code roughly separates them into small (blue), intermediate (green)

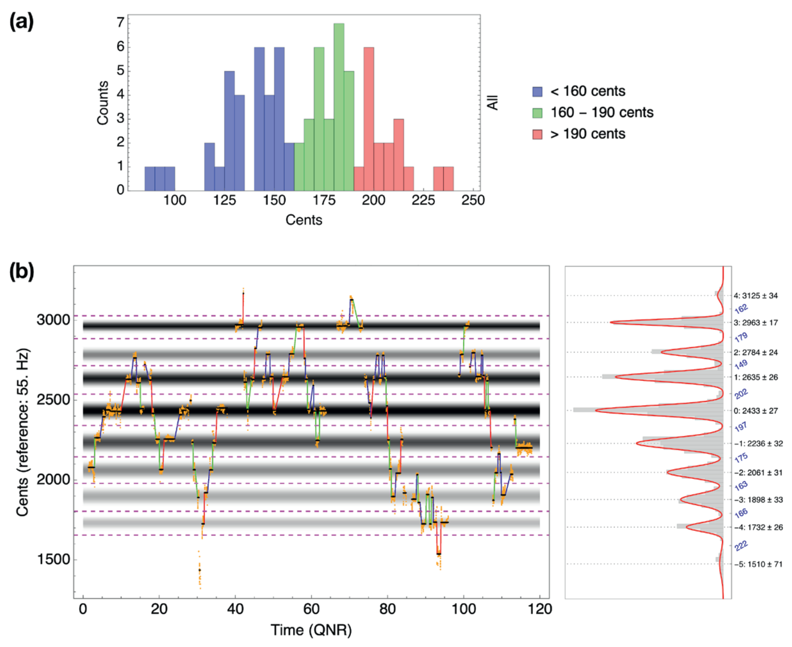

Figure 3: (a) Actual melodic step sizes between neighboring pitch groups in the top voice of chant GCH-ID 008 as they were sung by Artem Erkomaishvili. (b) Occurrence of the differently sized steps with respect to the pitch and note track. The grayscale in the left panel shows the probability density values of the Gaussian mixture distribution displayed in the right panel. The magenta horizontal dashed lines mark pitch group boundaries.

and large (red) ones. Figure 3(b) shows where in the chant the individual types occur. From this plot, it can be seen that there is no simple relation between the expected step sizes for melodic seconds (defined above as the pitch difference between the center pitch value of the pitch group from where the melodic step starts and the one where it ends) and the actual ones. The transition between a particular pair of pitch groups is sometimes realized by a small, sometimes by an intermediate, and sometimes by a large melodic step. A very similar observation was made by Arom, Fernando, and Marandola (2007) for the musical scales of the four-part a-cappella vocal music of the Bedzan Pygmies in Cameroon. They also observed relatively wide margins of realization for the scale-defining intervals as well as for the scale pitches. Their investigation, in contrast to our study, had the advantage that they could verbally interact with the musicians in the field. They could suggest a working hypothesis which could then be tested by the musicians and either verbally accepted or rejected. In this way, Arom, Fernando, and Marandola (2007) could identify the Bedzan's scale model as being dynamic based on a small set of rules, which act as constraints during the singing. In the present study, unfortunately, we can only evaluate the acoustic data of the recordings. The structure of the chants and the fact that the corpus contains a number of redundant melodic phrases, however, can partially compensate for this.

5.1.1 Section-Based Analysis

One structural property of the chants is that they are musically subdivided into individual sections, cf. Figure 1(b). These sections are usually separated by a rest and a melodic jump. The sizes of the melodic jumps are often considerably larger than the melodic step sizes within a section. It can be seen in Figure 4 that, for chant GCH-ID 008, the interval structure of the pitch distribution is visually similar for the individual sections. However, the relative heights of the individual peaks change across the section-based distributions. These heights correspond to the salience of the corresponding pitch group within a section. Although it is not visually apparent, the pitch distributions for the individual sections exhibit a slight pitch offset with respect to each other which can be determined by the position of the maximum values of their cross-correlation functions. For the chant group CG 008-011, the standard deviation (σ) of this pitch shift was calculated to be approximately 8 cents. This value is interpreted as a rough estimate for the unintended random pitch fluctuations in Artem Erkomaishvili's voice production. In Figure 3, it can be seen that on the level of the complete chant, there is no simple relation between the pitch differences between the center pitch values of neighboring pitch groups and the

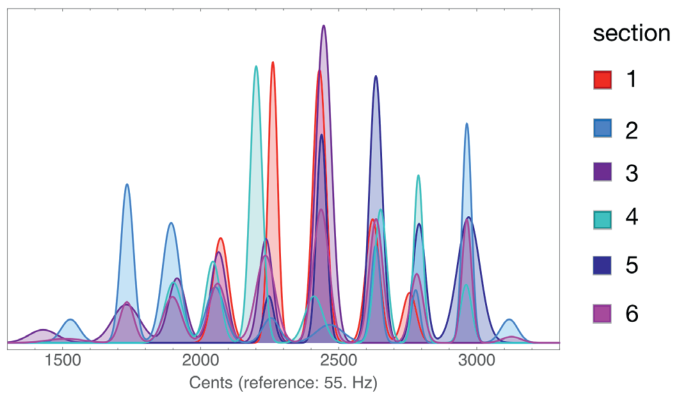

Figure 4: Section-based pitch distributions for chant GCH-ID 008. The sections are indicated by color code.

actual melodic intervals in a melody. Would this change if the analysis of the pitch distributions was performed on a section level?

Figure 5, which displays the section-based pitch distributions as density plots superimposed with the pitch and note track and the differently sized *melodic seconds* for chant GCH-ID 008, shows that this is obviously not the case. Between two neighboring pitch groups, e.g. the one with a center pitch close to 2400 cents and the one with a center pitch close to 2600 cents in the third section (for QNR 40–60) of Figure 5, one can observe melodic intervals from the small (green), the intermediate (blue), and the large (red) category. Hence, the differences between the center pitch values of neighboring pitch groups do not predict the actual melodic step sizes realized between them, at least not with any practically useful precision. This appears puzzling, since sound scales are often constructed simply by concatenating melodic intervals determined from a melodic sequence (see Tsereteli and Veshapidze 2014). This raises the question of what constrains the melodic step sizes between two pitch groups if not the difference between their center pitch values (the μ_k)? In other words, where is the scale?

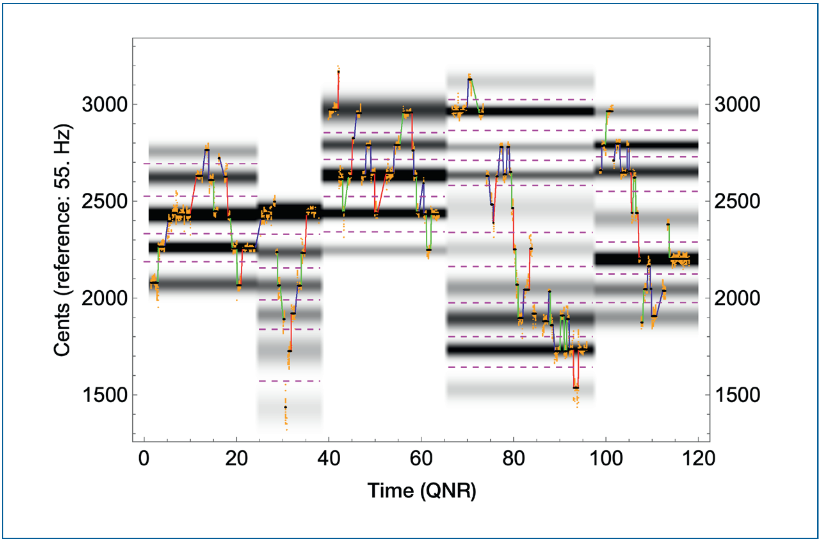

Figure 5: Differently sized melodic seconds with respect to the pitch and note track superimposed on the section-based pitch distributions for chant GCH-ID 008. The grayscale background shows the probability density values of the Gaussian mixture distributions for each of the sections. The magenta horizontal dashed lines mark the pitch group boundaries in each section.

A hypothesis which we tested was if it actually matters from where in a pitch group relative to the center pitch value μ_k a melodic step starts. Does it matter if the starting pitch of a melodic step is below, equal to, or above the center pitch value of the corresponding pitch group? In other words, if the pitch of the **starting** note in a melodic step under consideration deviated from the center pitch of the corresponding pitch group, would Artem Erkomaishvili unconsciously or consciously change the melodic step size of the **next** step to compensate for this deviation? The result of this test is shown in Figure 6. Each point in the diagram corresponds to one melodic step in the top voice of chant GCH-ID 008. The pitch group corresponding to the starting note under consideration is indicated by the color code explained in the legend to the right. The horizontal position of a dot displays the deviation of the pitch of the starting note (of a melodic step) from the center pitch value of its corresponding pitch group (μ_k). The vertical position of a dot corresponds to the extra step size (in addition to the difference between the center pitch value of target pitch group of the **next** melodic step and the center pitch value of the pitch group of the starting note).

Figure 6 clearly shows that there is a strong correlation between the extra step size of a melodic step and the pitch deviation of the starting note from the center pitch value of its corresponding pitch group. We interpret this as a clear proof that the melodic sound scale defined as the ordered set of pitch group center values of the whole pitch distribution indeed constrained Artem Erkomaishvili's melodic singing, although not in a rigid but a flexible way. The melodic sound scale for Artem Erkomaishvili does not define fixed melodic step sizes. It rather seems to act like a set of *attractors* around which the melody can more or less freely move, but to which it is always "pulled" back. One does not immediately see the presence of the melodic sound scale when looking only at a sequence of melodic steps, but it always acts as constraint during the melodic development.

The well-defined slope in Figure 6 (even for very small pitch deviations from the associated pitch group means) indicates that Artem Erkomaishvili seems to have been well aware of even small pitch deviations of a sung note from the central pitch value of the corresponding pitch group because he compensates for these deviations in subsequent melodic steps. His sensitivity for pitch deviations within a pitch group does not seem to have been lowered significantly, which could be interpreted as evidence against a strong perceptual magnet effect on his voice perception/production. Since we have no means to investigate this further, we can only state this here as an observation. The vertical scatter of the extra step sizes in Figure 6, which has a standard deviation of 16.5 cents, can be seen as a measure of precision with which the melody is following the scale pitches, for reasons which are addressed in the following.

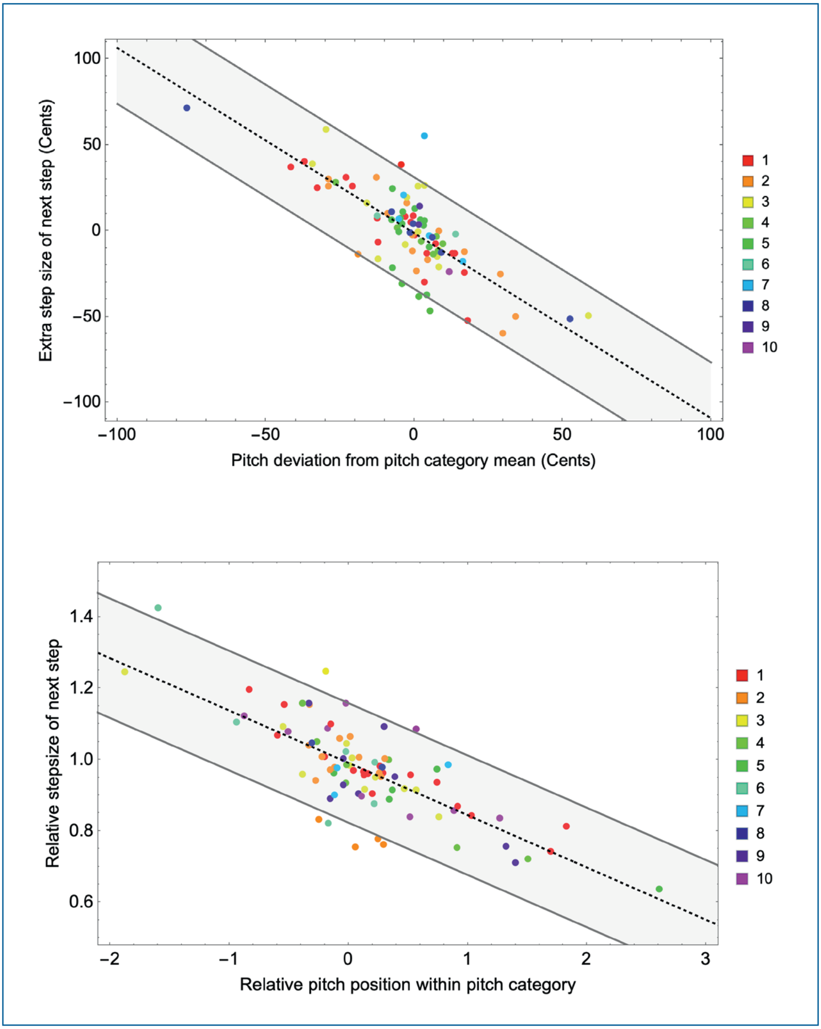

Figure 6: Distribution of extra step sizes of melodic seconds against the relative position of the previous note with respect to its corresponding pitch group center. The gray shaded area shows the linear regression line (dotted) $y = -1.5 - 1.1x \pm 33$ (2σ).

5.1.2 Intentional Pitch Bending

The gray shaded area in Figure 6 marks the region which, in case the scatter around the dotted regression line is Gaussian distributed, will contain roughly 95 percent of all extra step sizes. This scatter will include the scatter caused by perceptional effects discussed in Section 4.1.1, random effects as well as the effect of intentional pitch bending (using *shinpardi*). In the following, we estimate the order of magnitude of this effect by making use of the occurrence of a number of identical melodic phrases in the chant group CG 008-011.

Figure 7 shows the pitch trajectories of the melodic phrases which appear in identical form in all chants of CG 008-011. The one in the lower left panel appears only in chants with GCH-IDs 008, 009, and

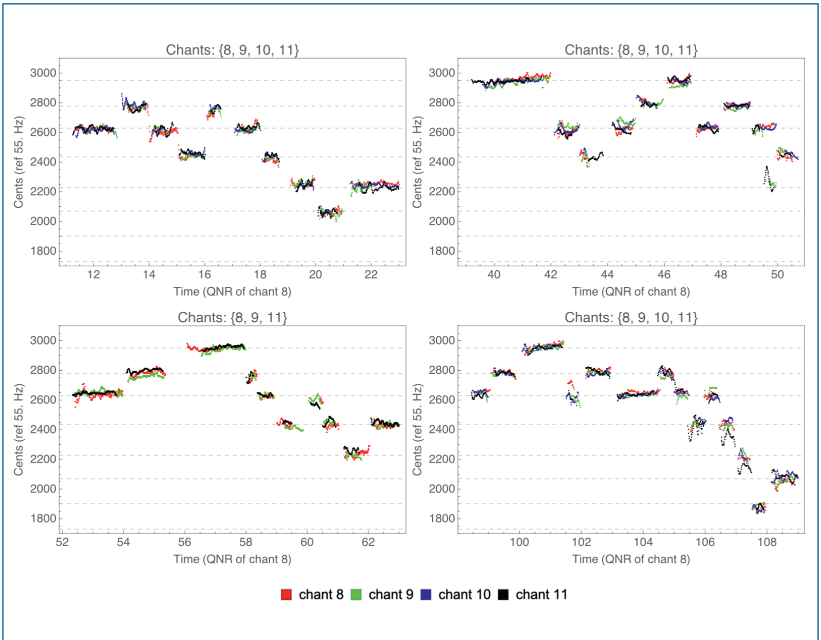

Figure 7: Pitch trajectories of the melodic phrases which appear in identical form in all chants of chant group CG 008-011.

011. The F0-trajectories have been realigned in time to the QNR grid of chant GCH-ID 008. For a detailed definition of this concept see Rosenzweig et al. (2020). In addition, the trajectories have been shifted in pitch such that the position of the maxima of the cross-correlation functions of each pitch distribution pair appears at position zero. The dotted horizontal lines in Figure 7 show the center values of the joint pitch distribution after pitch shifting. In other words, the dotted lines in Figure 7 represent the scale pitches of a joint scale model for all four chants.

In this plot it can be seen that – with very few exceptions – the trajectories show significant overlaps in average pitch although they correspond to independently made recordings and despite the fact that they can consistently deviate from the pitches of the average scale model. In order to obtain an estimate of the magnitude of this effect, we analyzed the first of the phrases in more detail. Figure 8 shows the corresponding score representation (a), the pitch trajectories superimposed by the note tracks (b), and the melodic steps between the notes as stem plots (c). As already indicated by Figure 7, the note events in the phrase are highly correlated for the four different chants. The gray horizontal lines in Figure 8(c) indicate the step sizes expected from the pitch distances of the scale pitches for the joint melodic scale model. In those cases, where the actual melodic steps do not closely cluster around the expected step sizes, the plot shows a systematic and consistent offset. This means that actual melodic steps tend to be either consistently smaller or larger than the expected value from the pitch group mean value differences and suggests that the deviation of the melodic step sizes from these expected values is a conscious choice of the singer.

Finally, we calculated the differences between the note pitches as predicted from the joint melodic scale model and the actual phrase notes. These values vary between 5 and 30 cents for the ten notes in the phrase. Although these values only represent a small number of notes, their magnitude demonstrates that the effect of intentional pitch bending is non-negligible and needs to be considered when trying to determine melodic scale models. The effect of intentional pitch bending (using *shinpardi*) could therefore also contribute significantly

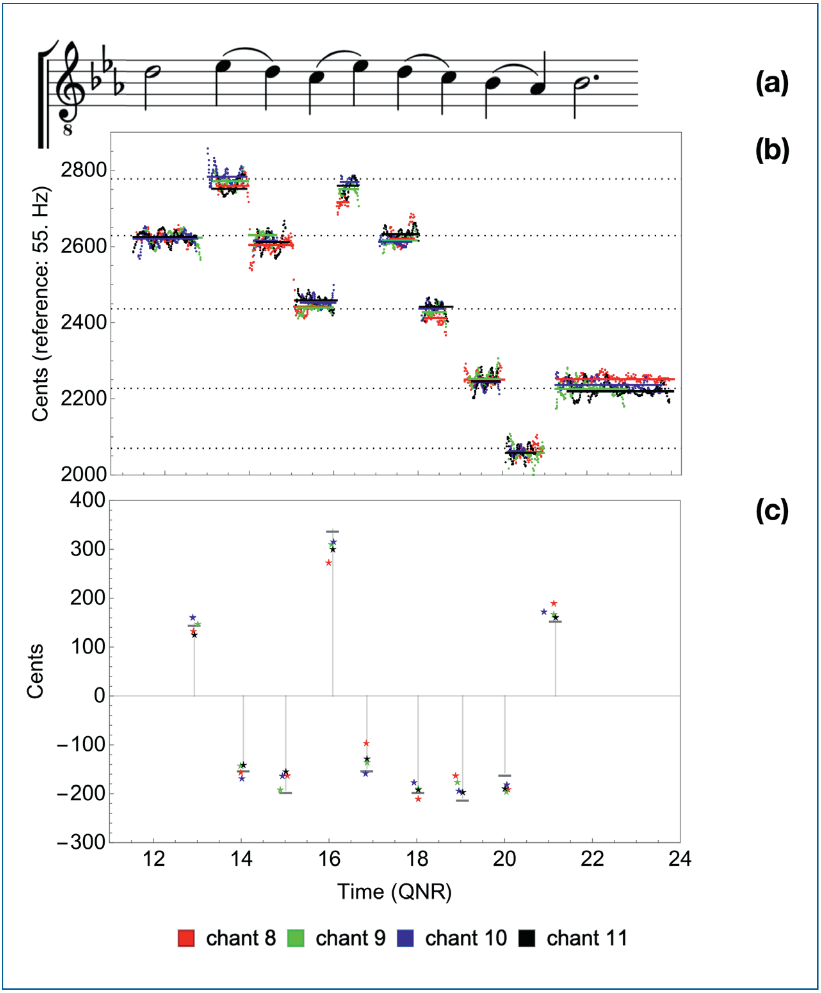

Figure 8: (a) Score representation. (b) Note and pitch trajectories. (c) Stem plots of the melodic steps of the first of the melodic phrases in Figure 7.

to the shape of the melodic step size distribution shown in Figure 3 as well as to the spread of the individual pitch groups in Figure 2.

It seems worth emphasizing that Figure 8 and Figure 6 illustrate two different mechanisms which jointly influence how Artem Erkomaishvili developed a melodic sequence. On the one hand, Figure 8 demonstrates that he did not rigidly adhere to fixed melodic steps as defined by the pitch differences between the scale pitches of the whole pitch distribution, but that he uses intentional pitch bending. On the other hand, Figure 6 clearly demonstrates that he is always aware of the position of a note which he sings with respect to the pitch of the corresponding pitch group center, which then influences his next melodic steps. As a consequence of this, individual melodic step sizes do not directly represent the intervals of the underlying melodic scale because melodic step sizes can be intentionally stretched or compressed as a matter of Artem Erkomaishvili's personal interpretation or they can be shortened or elongated by the mechanism documented in Figure 6.

If our assumption is correct, it is indeed the ordered set of pitch group center values of the whole pitch distribution (the μ_k values of the Gaussian mixture model) which defines the melodic sound scale. Melodies in this model are the result of a *deviation-compensation* mechanism, which clearly honors scales but still allows for a large degree of melodic flexibility.

5.2 Corpus-Level Analysis

In the following, we describe the results of the analysis of the top voice recordings for the complete Erkomaishvili dataset (see Section 3). The goal in this context is to derive a scale model which jointly represents the melodic sound scales of the complete dataset in a synoptic way. Towards this end, we first investigate to what degree the melodic top-voice scale models for the individual chants vary within the corpus. In this context, visually and acoustically processing of all audio tracks with the Tony software (Mauch et al. 2015), as we did for the chant group CG 008-011, turned out to be too labor-intensive. Therefore, for the corpus-level analysis, we determine the stable regions in each of the F0-trajectories using the morphological filtering approach described in Rosenzweig, Scherbaum, and Müller (2019).

Using spot checks, we validated that the differences between pitch distributions obtained from morphological filtering and those obtained using the Tony software are marginal for the present purpose. We are confident that for the corpus-level analysis, the gain in processing speed by orders of magnitude by far outweighs the lack of user-guided manual corrections.

The purpose of the analysis described in the present section is not to perform a detailed study of the properties of the individual pitch distributions of all chants, but to test if they share some common features (in a statistical sense) which can be detected computationally. Since Artem Erkomaishvili started each chant at a different pitch (depending on how the pitches of the chant would fit into his vocal range), the individual pitch distributions might be shifted in pitch by an unknown amount, even if the same melodic scale might be employed. Similar as in Section 5.1.1, we correct these differences by aligning the pitch distribution using cross-correlation analysis.

More specifically, we align all pitch distributions with respect to a reference chant (GCH-ID 010). However, the weights of the different Gaussian mixture elements, which reflect the amount of pitches belonging to a specific pitch group, are quite different. Since the cross-correlation function is quite sensitive to amplitudes and we are at this point only interested in the interval structure of the pitch distribution, we *whiten* the individual Gaussians such that after whitening they all have the same amplitude in their probability density functions (when having an equal number of mixture elements). In this way, we make sure that the alignment is only dependent on the interval structure of the distributions. The process of alignment and whitening of the Gaussian mixture distributions is illustrated in Figure 9 for the four chants of chant group CG 008-011, which were discussed in Section 5.1.

The results of whitening and subsequent alignment of the pitch distributions of all chants is shown in the top and bottom panel of Figure 10. One can see that a common structure becomes visible in

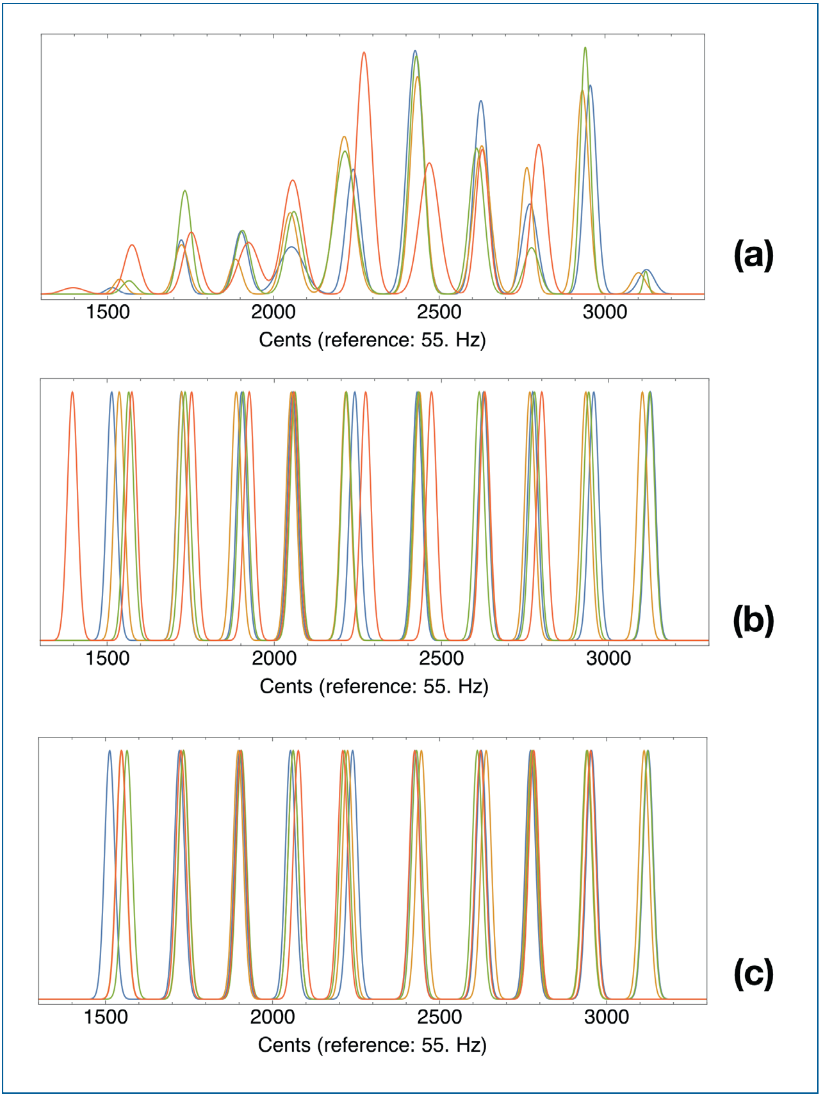

Figure 9: Illustration of the whitening and alignment process of the Gaussian mixture distributions for the chants of chant group CG 008-011. (a) Pitch distribution in raw form. (b) Pitch distribution after whitening. (c) Pitch distribution after whitening and alignment.

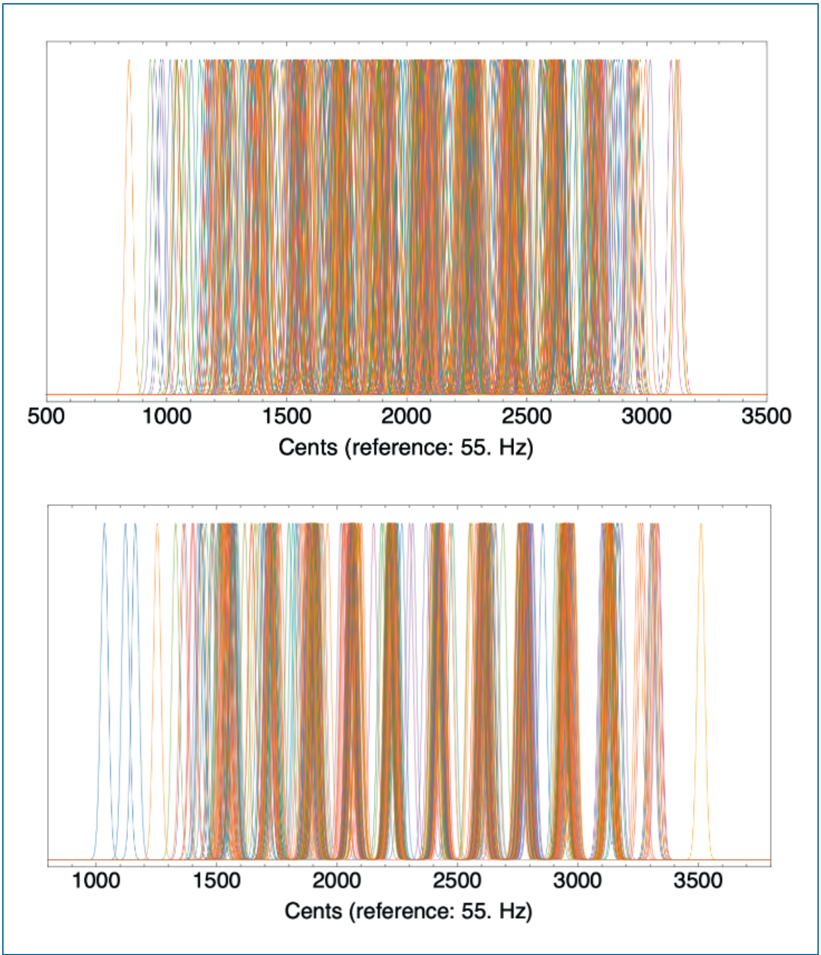

Figure 10: The top panel shows the whitened Gaussian mixture distributions for the top voices of all chants, while the bottom panel shows the distributions after whitening and alignment.

the superposition of all whitened and aligned distributions. This nicely illustrates that the shapes of the individual pitch distributions, which represent the melodic tuning models of the individual chants, are not completely different but can possibly – *cum grano salis* – be interpreted as variants of a common underlying structure. To determine the parameters of this underlying sound scale model, we take the collection of all the mean values of all Gaussian mixture model elements of the individual distributions from the bottom panel of Figure 10 as input data to calculate a histogram and the corresponding Gaussian mixture model. The resulting distribution is shown in Figure 11.

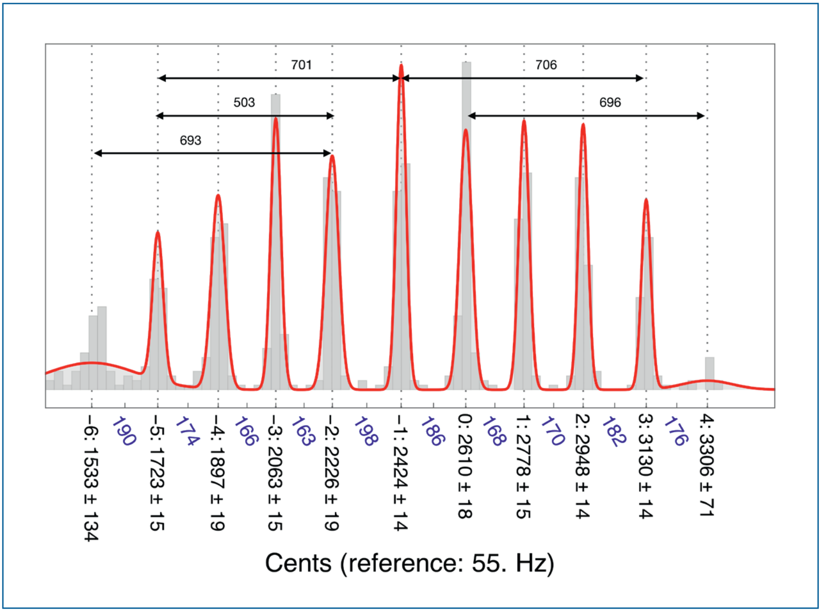

Figure 11: Histogram (gray shaded) and corresponding Gaussian mixture distribution (solid red line) representing the average pitch distribution for all top voices in the complete corpus. The black numbers before the colons show the pitch group indices (PGIs) with respect to the most salient pitch group (indexed as 0). The blue tilted numbers between the pitch groups show the intervals between neighboring pitch groups in cents.

Figure 11 can be interpreted to represent the properties of the interval structure of melodic sound scales of all the top voice tracks in an average sense, a kind of *executive summary*. The interval structures of chant-level pitch distributions can deviate from this corpus-level distribution in details which are averaged out by the processing above. The investigation of these details for all chants is beyond the scope of the present study. It can be noted that even at this coarse level, the interval structure of the average pitch distribution model for the top voice, which is numerically displayed in Table 1, shows musically very

Table 1: Pitch differences between all possible pitch group combinations in Figure 11. The individual table elements correspond to the combination of pitch groups for which the pitch group indices PGI*s* are given in the first column and the first row, respectively.

	−6	**−5**	**−4**	**−3**	**−2**	**−1**	**0**	**1**	**2**	**3**	**4**
−6	0	190	364	530	692	891	1077	1245	1415	1597	1773
−5	0	0	174	340	**503**	**701**	887	1055	1225	1406	1583
−4	0	0	0	166	328	527	713	881	1051	1233	1409
−3	0	0	0	0	162	361	547	715	885	1067	1243
−2	0	0	0	0	0	199	385	553	723	904	1080
−1	0	0	0	0	0	0	186	354	524	**706**	882
0	0	0	0	0	0	0	0	168	338	520	696
1	0	0	0	0	0	0	0	0	170	352	528
2	0	0	0	0	0	0	0	0	0	181	358
3	0	0	0	0	0	0	0	0	0	0	176

interesting features. For example, if we define a pure fourth as 498 ± 5 cents and a pure fifth as 702 ± 5 cents, then there are two pure fifths and one pure fourth visible in the corpus-level pitch distribution. As a side note, we point out that a feature, which we also observed quite prominently in individual pitch distributions, is the existence of rather large intervals (approx. 200 cents) between neighboring pitch groups close to the most salient one. Musicologically, this is a key finding, since it is the existence of this larger interval which is necessary to build the 1-4-5 structure, which has been described by many scholars (Aslanishvili 2010; Dimitri Araqishvili 2010; Chkhikvadze 2010; Jordania 2010) as being characteristic for traditional Georgian music. We find it extremely interesting to see that the 1-4-5 structure, which is typical for the harmonic structure of traditional Georgian music, is clearly visible as a melodic summary feature of the pitch distribution of all top voices.

6 Multi-Voice Results

The most characteristic musical feature of traditional Georgian vocal music is its polyphonic or multi-voice character. In the following, we determine the pitch and harmonic interval distributions. Instead of considering only the top voice (as in Section 5), we now consider all three voices (top, middle, bass). We extend our analysis in three ways. First, in Section 6.1, we consider pitch distributions for all three voices. Second, in Section 6.2, we consider harmonic intervals by comparing the voices in a pairwise fashion. In Section 6.3, as extension of the discussion in Section 5.1.1, we analyse the relationship between the melodic progression and the multi-voice pitch distributions. Subsequently, we derive synoptic pitch and interval distributions for all voices and all chants and discuss their properties.

6.1 Pitch Distributions

For the determination of the pitch distributions in the multi-voice case, we reduce the stable regions of the raw F0-trajectories of the individual voices to those regions which appear concomitantly in all three voices. This ensures that the F0-values for the analysis of the melodic and the harmonic properties are identical.

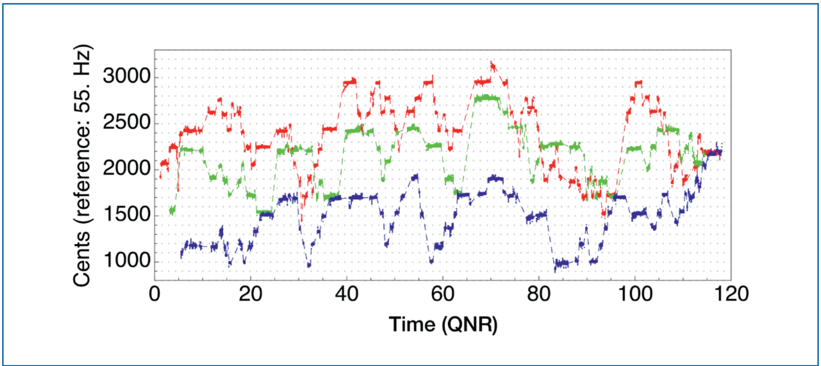

Figure 12: F0-trajectories for all F0-values in stable segments which appear concomitantly in all three voices of chant GCH-ID 008.

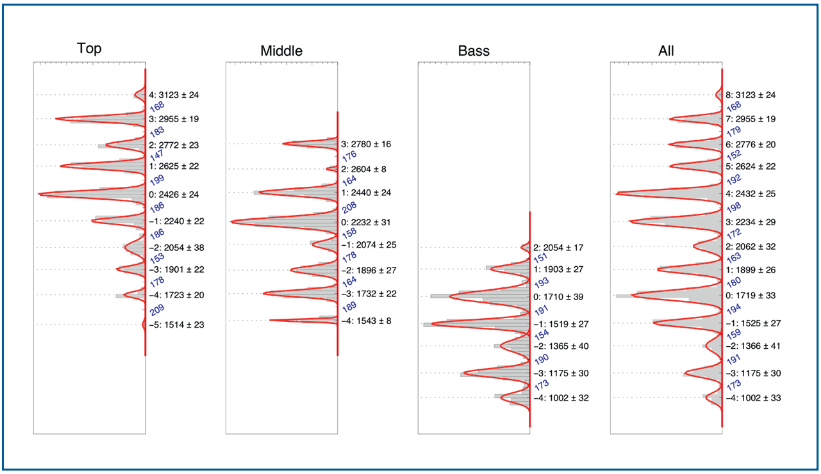

Figure 13: Gaussian mixture models of the pitch distributions for the three voices (top, middle, bass) as well as for all voices jointly, using chant GCH-ID 008 as example.

The corresponding F0-trajectories superimposed by the corresponding note objects for chant GCH-ID 008 are shown in Figure 12. Figure 13 displays the corresponding Gaussian mixture models of the pitch distributions for the individual voices as well as for all the F0-values in all voices jointly. The pitch group mean values are given in cents with reference to 55 Hz.

The center pitch values for the comparable pitch groups vary between the individual voices in the order of 10–20 cents. We can also observe that the pitch groups' standard deviations of the bass voice are larger than for the top and middle voices.

6.2 Harmonic Interval Distributions

Figure 14 shows the GMMs for the harmonic interval distributions for the different voice combinations of chant GCH-ID 008, as well as for all voice combinations jointly. For the explanation of the labeling of the

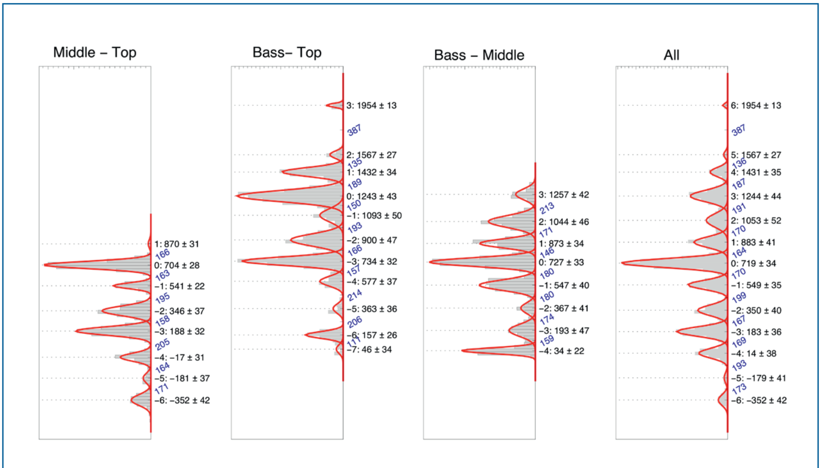

Figure 14: Gaussian mixture models of the harmonic interval distributions for the different voice combinations of the individual voices of chant GCH-ID 008, as well as for all voice combinations jointly. For the explanation of the labeling of the individual Gaussian mixture elements see the figure caption of Figure 2.

individual Gaussian mixture elements, see the caption of Figure 2. It can be noted that the fifth is the most frequent harmonic interval in all voice combinations. Its tuning, however, is not always the same.

It appears as justly tuned (close to 702 cents) in the middle-top voice combinations and more than 20 cents larger in combinations where the bass voice is involved (734 ± 32 cents in bass-top and 727 ± 33 cents in bass-middle). The octaves appear stretched by 40 to 50 cents for all voice combinations where they appear in chant GCH-ID 008[4] (1243 ± 43 cents in bass-top and 1257 ± 42 cents in bass-middle).

6.3 Melodic Progression and Pitch Distributions

In the following, we investigate whether the melodic progression for the middle and bass voices is equally well described by its relation to the multi-voice scale model as it is for the top voice case shown in Figure 6. To this end, similar to Section 5.1.1, we again calculate the distribution of extra step sizes of *melodic seconds* against the relative position of the previous note with respect to its corresponding pitch group center for the bass voice (Figure 16, bottom panel), but this time for the section-level pitch distribution calculated from the pitches of all voices shown in Figure 15.

In Figure 15, the pitch distributions are determined for each section in chant GCH-ID 008 and the corresponding mixture models are displayed as grayscale image in the background. Figure 15 illustrates – in an overall sense – the tuning of this chant jointly for all the individual voices. From the grayscale distributions, we see that the changes of the pitch group boundaries for different sections in the case of chant GCH-ID 008 are rather small and restricted to a few cases (for example

4 The Gaussian mixture models for the pitch distributions for all voices and for the intervals for all voice combinations have been calculated for all chants. Their in-depth discussion, however, is beyond the scope of the present paper, but they are made available on request from the first author.

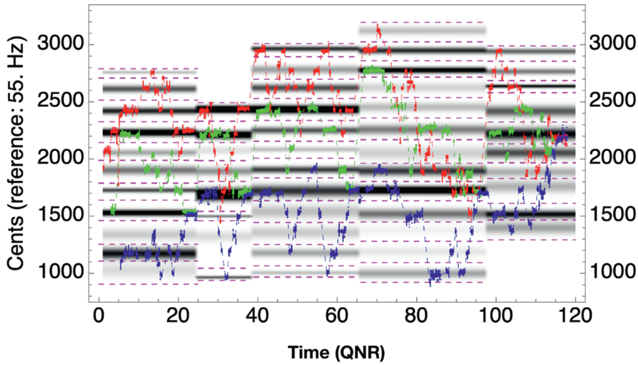

Figure 15: F0-trajectories and note objects from Figure 12 superimposed on the segment GMMs for all voices of chant GCH-ID 008. The grayscale background shows the probability density values of the Gaussian mixture distributions for each of the segments. The magenta horizontal dashed lines mark the pitch group boundaries in each segment.

in the 2nd section). Visually more obvious is the change of the center values for identical pitch groups across section boundaries. This corresponds to changes in the perceptional strength of a pitch group for different sections. The darker the grayscale value for a pitch group, the more frequent are pitches from this pitch group in a section.

For comparison, the top panel in Figure 16 shows the case of the top voice again (with the scale model calculated from the top voice), displayed in the same plot range as the bass voice. The gray shaded areas show the linear regression lines (dotted) $y = -1.5 - 1.1x \pm 33$ (2σ) and $y = -3.0 - 1.2x \pm 64$ (2σ) for the top and the bass voice, respectively.

The slopes of the regression lines are essentially identical and the offsets for x close to zero as well (1.5 and 3 cents, respectively), but the σ value describing the presumably random fluctuation around the regression line in the case of the multi-voice scale model is twice as large as for the top voice model. This means that, although the correlation between the size of the extra step sizes of the *melodic seconds*

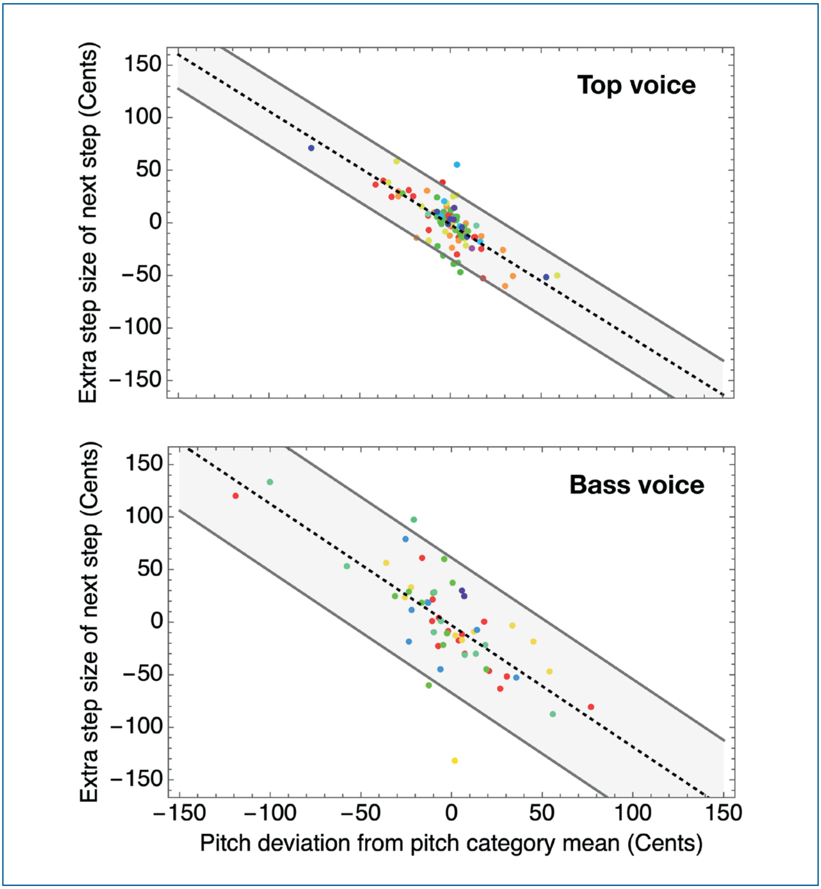

Figure 16: Distribution of extra step sizes of melodic seconds against the relative position of the previous note with respect to its corresponding pitch group center for the top and the bass voice. The data for the top voice are identical to the ones in Figure 6.

and the relative position of the previous note with respect to its corresponding pitch group center is as clear for the bass voice as it is for the top voice. The absolute step sizes of melodic step sizes between neighboring pitch groups can only be predicted at a much lower precision. One possible explanation for this behavior could be that it is an effect of Artem Erkomaishvili adjusting his middle- and/or bass voice pitches in order to achieve stable harmonic intervals at certain positions in a chant. This phenomenon has already been mentioned as qualitative observation in one of the first non-Georgian treatises on traditional Georgian vocal music (Nadel 1933). In Section 7, we address this question quantitatively.

6.4 Synoptical Models

One of the central questions of the current study is whether Artem Erkomaishvili's recordings preserve a characteristic tuning structure which can be extracted from the audio data. Motivated by the results shown in Figure 11 for the top voice, a similar approach was chosen for the polyphonic analysis. The resulting distributions (now for all voices) are shown in Figure 13 and Figure 14 in the far right-hand tracks. The resulting synoptic distributions for pitch and harmonic intervals based on all voices in the corpus are displayed in Figure 17 and Figure 18, respectively.

While individual chants differ in details, the structures, which are visible in the distributions computed on the whole corpus, reflect Artem Erkomaishvili's tuning in a condensed way.

The pitch differences between all possible pitch group combinations in Figure 17 are listed in Table 2. The individual table elements correspond to the combination of pitch groups for which the PGIs are given in the first column and the first row, respectively. What can be noted in Table 2 is that a number of pitch group combinations exhibit center pitch differences (of their μ_k values) which are close to a pure fifth (702 cents). In addition, the center pitch differences between a pitch group and its next but one is always roughly a *neutral third* (between 300 and 400 cents).

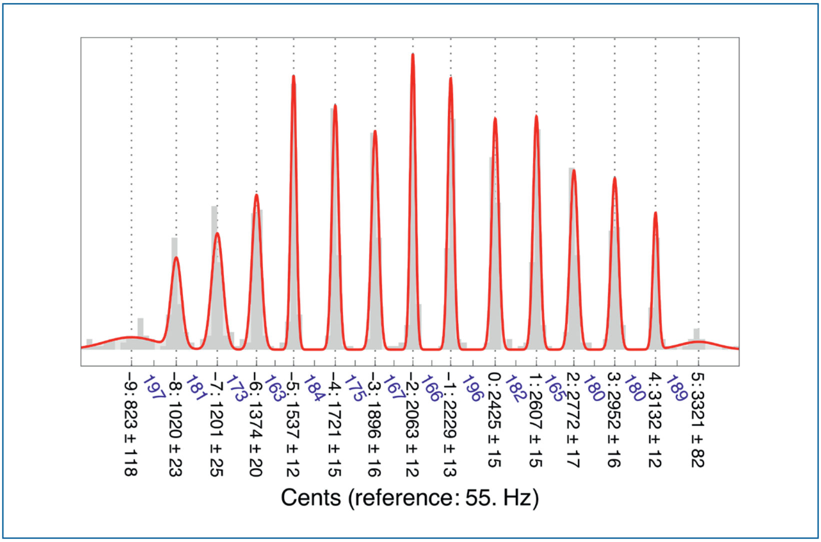

Figure 17: Corpus-level pitch distribution (for all voices and all chants).

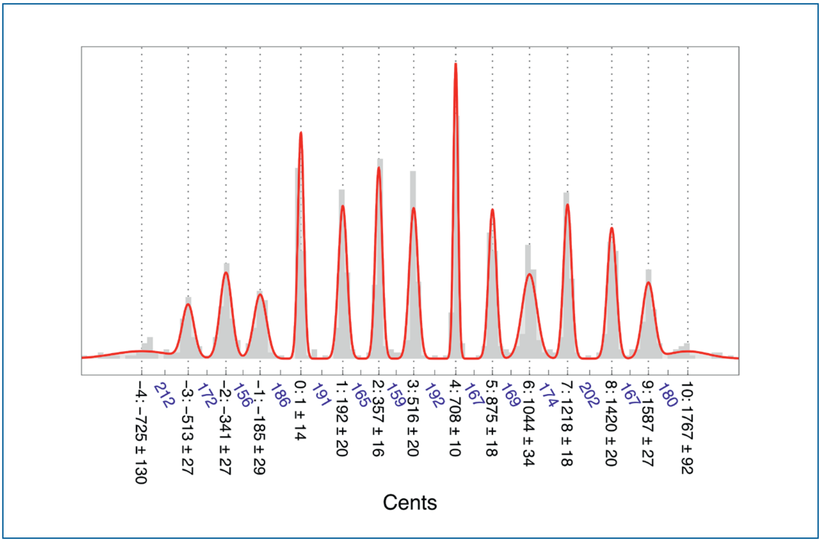

Figure 18: Corpus-level harmonic interval distribution (for all voices and all chants).

Table 2: Center pitch differences between all possible pitch group combinations in Figure 17. The individual table elements correspond to combinations of pitch groups as defined by the PGIs given in the first column and the first row, respectively.

	−9	**−8**	**−7**	**−6**	**−5**	**−4**	**−3**	**−2**	**1**	**0**	**1**	**2**	**3**	**4**	**5**
−9	0	197	378	551	**714**	897	1073	1239	1406	1602	1784	1949	2129	2308	2498
−8	0	0	181	353	517	**700**	876	1042	1209	1405	1587	1752	1932	2111	2301
−7	0	0	0	173	336	520	**695**	862	1028	1225	1406	1571	1751	1931	2121
−6	0	0	0	0	163	347	522	**689**	856	1052	1233	1398	1578	1758	1948
−5	0	0	0	0	0	184	359	525	**692**	888	1070	1235	1415	1595	1784
−4	0	0	0	0	0	0	176	342	509	**705**	886	1052	1231	1411	1601
−3	0	0	0	0	0	0	0	166	333	529	**711**	876	1056	1235	1425
−2	0	0	0	0	0	0	0	0	167	363	544	**710**	889	1069	1259
1	0	0	0	0	0	0	0	0	0	196	378	543	**722**	902	1092
0	0	0	0	0	0	0	0	0	0	0	181	347	526	**706**	896
1	0	0	0	0	0	0	0	0	0	0	0	165	345	525	**715**
2	0	0	0	0	0	0	0	0	0	0	0	0	180	359	549
3	0	0	0	0	0	0	0	0	0	0	0	0	0	180	370
4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	190

The synoptic harmonic interval distribution in Figure 18 shows that the most common intervals are more or less justly tuned fifths (708 cents), followed by unisons (1 cent), neutral thirds (341 cents), sharp fourths (516 cents), stretched octaves (1218) and ninth (1420). It can also be noted that the standard deviation for the harmonic seventh, which peaks at 1044 cents is much wider than for the other intervals, indicating that sevenths are realized in a variety of ways in the chants.

7 Microtonality

In this section, we discuss some observations regarding microtonal aspects of the Erkomaishvili dataset. We start with the discussion of the melodic step size distribution for the complete corpus in Section 7.1. This will illustrate, yet from a different perspective than in the previous sections, that Artem Erkomaishvili varied his melodic step sizes on a microtonal level by freely deviating from the expected melodic intervals defined by the scale pitches. In Section 7.2, we discuss the mechanism of harmonic intonation adjustment, which could explain one of the reasons for these microtonal melodic step size variations.

7.1 Corpus-Level Melodic Step Size Distribution

One of the most striking characteristics of the pitch distributions presented is that the smallest inter-cluster distances, or next neighbor intervals, in Figure 2 and Figure 17 are in the order of 150 cents. Semitone-sized differences (e.g., 100 cents) between neighboring pitch groups in the pitch distributions are essentially absent, which might at first glance suggest that this would also be reflected somehow in the melodic step size distribution. This, however, is not the case, as can be seen in the synoptic melodic step size histogram which is shown for all chants and all voices in Figure 19.

The figure shows that, although the most frequent melodic step size lies between 170 and 180 cents (cf. the red circles in Figure 19),

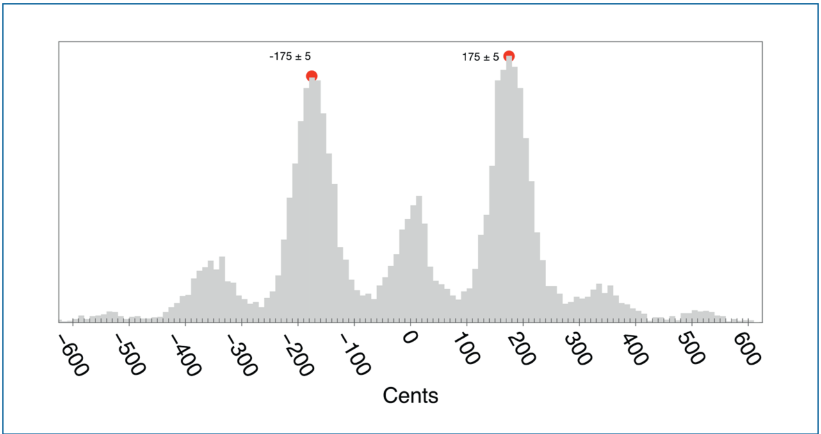

Figure 19: All melodic step sizes in all voices and all chants. The two red circles mark the bins with the most frequently occurring positive and negative step sizes, respectively.

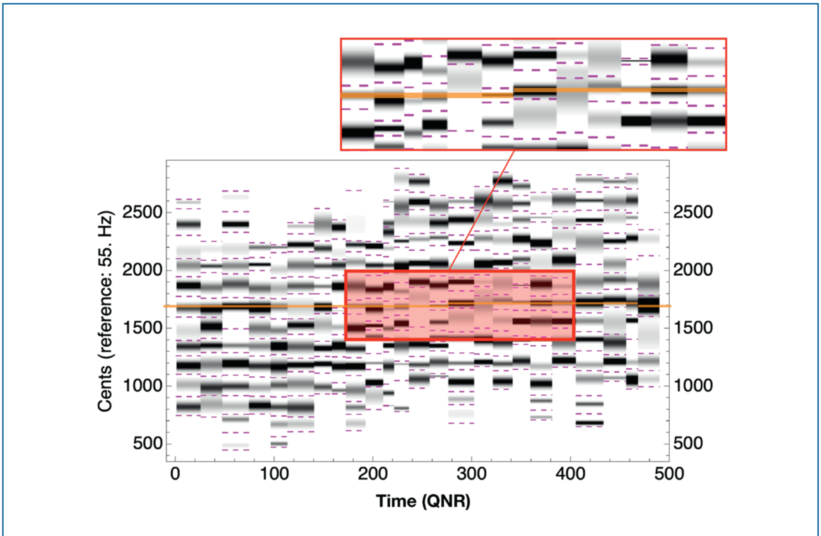

Figure 20: Probability density values of the Gaussian mixture distributions for each of the 24 sections of chant GCH-ID 026 (grayscale background), superimposed with the pitch group boundaries in each section shown by the dashed magenta lines. The orange line indicates the progression of center pitch of a pitch group.

essentially all values between zero and ±700 cents have been realized with different frequencies of occurrence somewhere in the dataset.

A particularly impressive example of microtonal pitch changes is chant GCH-ID 026, which contains a total of 24 different sections between which the whole tuning system varies on a microtonal level. Figure 20 shows the probability density values of the Gaussian mixture distributions for each of the sections superimposed with the pitch group boundaries in each section shown by the dashed magenta lines. The orange shaded thick lines of Figure 20 for example mark a pitch group for which the center pitch stays more or less constant for the first 14 segments, whereas it is shifted upwards by 30–40 cents from segment 15 on (see zoomed-in area, shift at QNR 280). This level is kept until the next to last segment, were the center pitch is going back to essentially the value of the first section. This example illustrates not only the variability of the tuning systems which Artem Erkomaishvili sometimes employs, but also that he still had excellent pitch control.

Another aspect of microtonal pitch shifts is related to how the harmonic structure of a song affects how notes in a melodic sequence are sung. On the one hand, harmonic interval perception is known to influence the fine-tuning of voices, which can lead to unintended pitch shifts of a whole ensemble (see Howard 2007; Mauch, Frieler, and Dixon 2014). On the other hand, harmonically driven intonation adjustments can also be intended in order to achieve certain harmonic intervals with particular precision. This phenomenon is frequently occurring in traditional Georgian vocal music and is often referred to as as "vertical musical thinking" (cf. Nadel 1933; Chokhonelidze 2010).

7.2 Dynamic Adjustments

In the following, we discuss to what degree intonation adjustments are measurable in the Erkomaishvili dataset. In particular, we focus on a phenomenon that we refer to as *dynamic intonation adjustment,* which expresses that two voices move in a correlated fashion to achieve a particular harmonic interval. When singing the middle voice, for example, Artem Erkomaishvili was hearing the pre-recorded top voice.

Therefore, his top voice pitch trajectory was already fixed. Hence, all we need to measure is if the pitch trajectory of the top voice influenced his intonation of the middle voice and similarly his bass voice. Evidence that these effects might be observable in the Erkomaishvili dataset was reported by Scherbaum, Müller, and Rosenzweig (2017a), who observed an overall correlation between concomitant pitch pair values for different voice combinations.

Our approach to measure dynamic intonation adjustment is based on the analysis of pitch fluctuations of short pitch trajectories. Our hypothesis is that dynamic intonation adjustment introduces statistical dependencies between pitch trajectories of two voices, which can be mathematically quantified through the analysis of the variances of the individual pitch trajectories and their differences. The theoretical basis is given by the fact that the difference of two uncorrelated Gaussian random variables (RV) is again a Gaussian RV with the mean being the difference of the means and the variance being the sum of the variances (Stirzaker 1999). In case of a correlation between the two RVs (as it is expected in the case of pitch tracks where both singers try to maintain a particular interval despite fluctuations of the individual voices), the variance of the interval trajectory will be less than the sum of the individual variances. Figure 21 shows an example for such a case, where the variance of the fluctuations of the interval trajectory for a single note of 2.5 sec duration shown in Figure 21(b) is 0.36 of the sum of the variances of the two individual trajectories, shown in Figure 21(a). Figure 21(c) shows the correlation between the pitch values for the two trajectories.

It needs to be emphasized that the approach described above only captures intonation adjustments in which two voices fluctuate and continuously move in a correlated way. Trying to correlate the fluctuations of the pitches of one voice with another one can also be seen as an indirect indicator of which harmonic intervals singers are trying to intonate as purely as possible, in other words, which intervals they "care about" harmonically. Figure 22 shows the number of occurrences for all those dynamically adjusted harmonic intervals in the complete dataset which consist of notes longer than 0.5 seconds and for which the variance reduction of the interval variance was at least 50 percent.

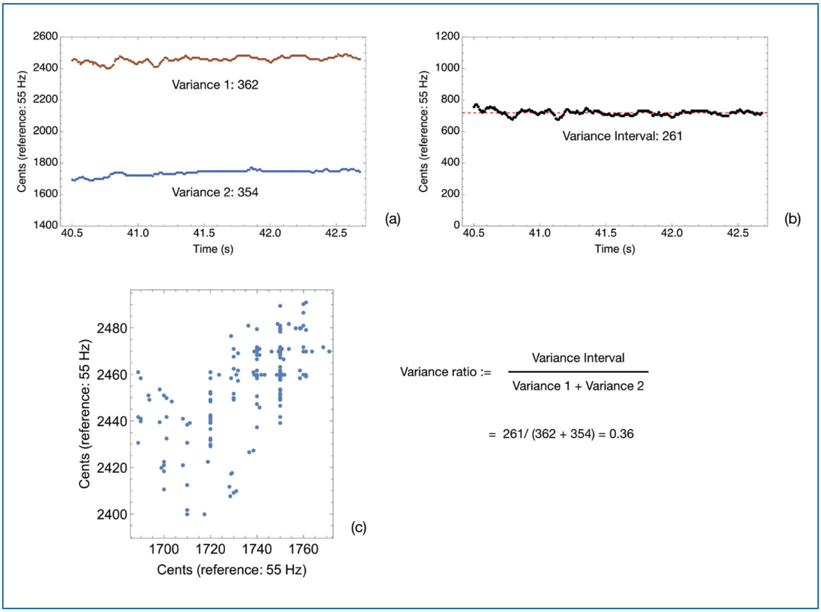

Figure 21: Example for dynamic intonation adjustment from chant GCH-ID 010. (a) Pitch trajectories for two notes building a harmonic interval of a fifth. (b) Corresponding harmonic (signed) interval trajectory, which is defined as the difference of the 'higher voice' in the sequence top, middle, bass, and the 'lower voice'. (c) Correlation between the pitch values of the two trajectories.

Table 3 shows the corresponding mean values (μ) and standard deviations (σ) per musical interval in cents.

The intervals which are most frequently dynamically adjusted in intonation are the fifth, octave, and unison, followed by the ninth and the fourth. These are also the intervals which in Table 3 are most precisely defined (smallest standard deviation). Examples for adjusted 3rds, 6ths, and 7ths are also visible in Figure 22 but their interval values (column 2 in Table 3) are much less well defined as can be seen from their increased standard deviations (column 3). Since the adjusted intervals in Table 3 result from an intentional process, and not just from averaging, their numerical values (column 2) might represent the

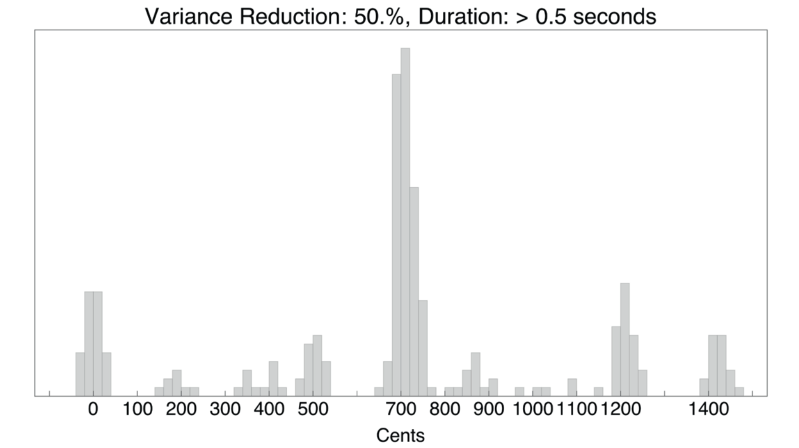

Figure 22: Histogram of the number of occurrences for those dynamically adjusted harmonic intervals in the complete dataset which consist of notes longer than 0.5 seconds and for which the variance reduction of the interval variance was at least 50 percent.

Table 3: Mean values (μ) and standard deviations (σ) per musical interval based on histogram in Figure 22 in cents.

	(μ)	(σ)
Unison	−2	18
2nd	190	22
3rd	382	34
4th	504	21
5th	707	18
6th	866	29
7th	1042	57
8ve	1211	16
9th	1420	14

harmonic tuning system which Artem Erkomaishvili was aiming at as closely as possible.

8 Concluding Discussion

With his performances of more than one hundred liturgical chants recorded at the Tbilisi State Conservatory in 1966, Artem Erkomaishvili has left a unique documentation of past performance practices. The fact that he sang the three voices of each chant in succession – the second and third voice against his own pre-recorded voice(s) – makes this corpus, in particular in its curated version by Rosenzweig et al. (2020), an extremely valuable resource for the analysis of traditional Georgian music. This dataset is the oldest corpus of Georgian chants from which the time synchronous F0-trajectories for all three voices have been reliably determined (Müller et al. 2017). Since chanting was prohibited during the Soviet period in Georgia, it is also **the only one** from this time, which further underlines its importance for research. In the present study, we made heavy use of the metadata of the curated dataset (Rosenzweig et al., 2019), all of which are publically available. Without the beat annotations and the possibility to easily align recordings computationally to positions in the scores, it would have hardly been possible to investigate the effect of intended pitch bending by comparing identical melodic phrases in different chants and to study the effects of dynamic intonation adjustments at a note level. We have also carefully documented all the processing steps to make our work transparent and reproducible, and we are happy to make the results for the individual chants available on request from the first author. As a consequence, we hope that the provision of all the data and methods to the scientific community, will stimulate further research on the Erkomaishvili dataset, for example on individual chants.

With the present study, we demonstrate that a lot can be learned about the tuning system(s) and the intonation practice of Artem Erkomaishvili from the Tbilisi State Conservatory recordings of 1966 which have wider implications for the discourse on the analysis of tradition Georgian vocal music. As one of our major contributions, we have

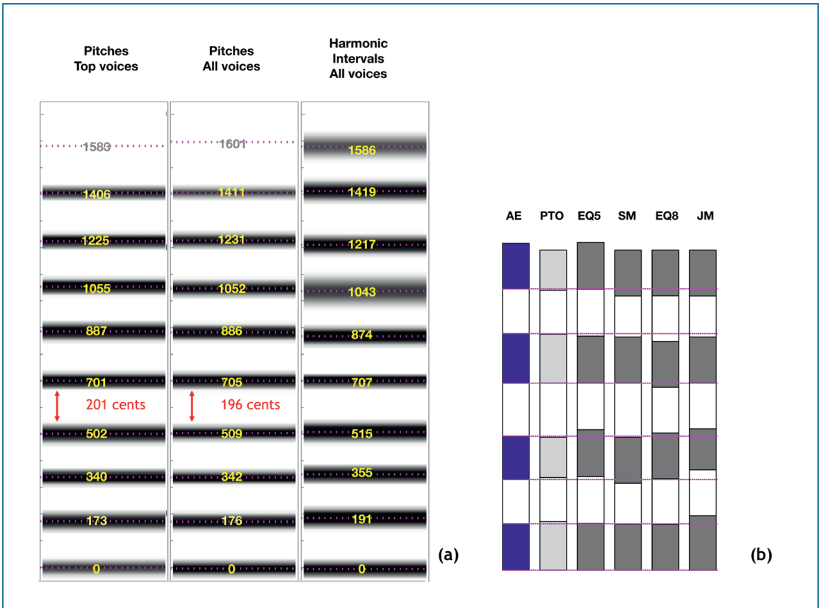

Figure 23: Synoptic scale models for the Erkomaishvili dataset. (a) Our three developed models. (b) Comparison of the synoptic top voice model (AE) with five selected scale models. The scales color-coded in dark gray correspond to scale models from the literature; the one in light gray to a scale model from the Scala library.

derived three synoptic scale models for the **complete dataset**, two melodic ones (one for the solo top voices and one for all voices jointly), and one harmonic model for all concomitant intervals in the dataset, see Figure 23(a). This is unprecedented in the context of the discourse of the Georgian tuning systems. Despite the fact that no data were excluded from the analysis, the synoptic models, which represent the complete dataset in a comprehensive way, show remarkable clear and stable features. The most persistent observation in essentially all chants is that their pitch group distributions contain two rather large intervals (approx. 200 cents) close to the most salient pitch group. Musically, this large interval seems to serve a very important purpose because it is an indispensable condition to have concomitant fourths and fifths (separated by 200 cents), which are the core of the 1-4-5 chord so common in traditional Georgian music. Therefore, in Figure 23(a) we have aligned the melodic scales to the scale pitches of the pitch group four degrees below the most salient one. This corresponds to the assumption that the most salient pitch group corresponds to scale degree five. As a consequence, in both the melodic and the harmonic models, the fourth and the fifth are close to just tuning.

Overall, melodic and harmonic scales are fairly similar, except for the first two scale degrees. In contrast to earlier, preliminary work (Scherbaum, Müller, and Rosenzweig 2017a), in the present study, we see clear evidence for enlarged octaves. It is known that exact octaves with a frequency ratio of 2:1 are generally perceived as too small (Burns 1999). This phenomenon, known as *octave stretching,* has been discussed in several papers (for references see Burns 1999), without a consensus being reached on the cause. Since octave stretching is a perceptional phenomenon and not a scale design characteristic, in the following comparison with other scale models, which partially have been derived from purely theoretical considerations or from instruments, we are not including the numerical value for scale degree eight.

As an add-on to the Scala software package, Op de Coul (2007) developed a large library of scales (more than 5000) which he has made freely available (http://www.huygens-fokker.org/docs/scales.zip). Out of these, we have extracted all the heptatonic ones (more than 700) and compared them quantitatively with the melodic sound scale derived from the top voice pitches. As distance metric, we used the Euclidean distance between the vectors constructed from the first seven scale degrees (ignoring the octave for reasons explained above). In this comparison, the results of which are displayed in Figure 23(b), we also included the scale models which have been proposed by Erkvanidze (2016), Tsereteli and Veshapidze (2014), and Gelzer (2002). Note that for better visual comparability of the interval structure of the individual scales, every second interval is color-coded. The scales are arranged from left to right according to their closeness to the top voice scale of Artem Erkomaishvili (color-coded in blue), which is labeled AE. The best fitting scale which nearly perfectly matches Artem Erkomaishvili's top voice scale (except for the octave, which was excluded from the comparison for reasons explained above) is what is referred to in the scale database as *inverse of Ptolemy's Equable Diatonic, 11-limit superparticular.* It belongs to the class of Ptolemy's mixed scales which according to Chalmers 1993 (page 109) were in common use by players of the lyra and kithara in Alexandria in the second century AD. Following in closeness is the scale labeled EQ5 which is derived by dividing a fifth into 4 equally sized intervals and which was suggested by Gelzer (2002). This scale, however, does not reproduce the pure fourth and fifth which seem to be a very clear characteristic feature of Artem Erkomaishvili's top voice scale. The next best fit is achieved by Malkhaz Erkvanidze's split mode (SM) model (Erkvanidze 2016), which fairly well reproduces the first six scale degrees, but not the seventh. It is followed in fit by the equidistance model of Tsereteli and Veshapidze (2014), labeled EQ8, and the joined mode model labeled JM of Erkvanidze (Erkvanidze 2016). None of these, however, reproduces the pure fourth and fifth of Artem Erkomaishvili's top voice sound scale. In terms of reproducing the pure fourth and fifth and providing a good overall fit to Artem Erkomaishvili's top voices' sound scale model, only Ptolemy's model and Erkvanidze's split mode model match well. Both are disjunct tetrachordal scale models.

Although Erkvanidze's split mode scale model comes fairly close to matching the first six degrees of Artem Erkomaishvili's top voice sound scale, the results of our analysis do not support Erkvanidze's assumption that the melodic steps of the old masters accurately follow the systemic scale with precision of a few cents (cf. Erkvanidze 2016). Instead, the tuning system used by Artem Erkomaishvili in his recordings from the year 1966 is clearly not a fixed and rigid one in which every single melodic step would have to correspond to an interval between one of the scale-degree-defining pitches. For example, Figure 5 unambiguously demonstrates that melodic step sizes vary freely even between the same pitch group pairs. What we see in our study is that Artem Erkomaishvili is well aware of his position with respect to the scale pitches at a very high precision, but exercises considerable freedom in stretching or compressing the melodic step sizes. We see the discovery and quantitative description of this *'deviation-compensation mechanism'* as another new and important finding of our work.

This mechanism also suggests an answer to the question of whether the determined melodic sound scale models can be interpreted beyond their descriptive value? In other words, to the question of whether there is any evidence that the scales constructed from the center pitches of the Gaussian mixture models (the μ_k) have any normative meaning in Artem Erkomaishvili's singing? If Westman (2002) was right in his scepticism regarding the principal usefulness of the scale concept, pitch intervals might simply be learned as elements of melodies without requiring an underlying normative scale model. In this case, the determined sound scale models would be purely descriptive. A similar view seems to be expressed in the famous quotation of C. Hubert H. Parry: "It is advisable to guard at the outset against the familiar misconception that scales are made first and music afterwards" (as quoted in Hornbostel 1913). The observed *'deviation-compensation mechanism'* clearly suggests that the scale pitches act in a normative sense as adaptive constraints for the choice of Artem Erkomaishvili's melodic progression. Otherwise, we should not observe a dependency of the melodic step size of the next melodic step on the relative position of the note pitch within the actual pitch category.

The *'harmonic intonation adjustment'* mechanism is yet another new contribution of our study. It has only qualitatively been described for Georgian music (e.g. Nadel, 1933) before, and is the final link to understand how the music 'works'.

Our results regarding the microtonal characteristics of the dataset and the dynamic intonation adjustments are not really unexpected. They underline again that Artem Erkomaishvili was an exceptional singer who, when singing the middle and bass voice, was able to instantaneously adjust his intonation to achieve desired harmonic intervals at a fairly high precision, judging by the amount of variance reduction in the interval traces which he was able to achieve.

On the methodological side, we have also entered new territory. Our study introduces a new approach to generate tuning models which are consistent with the melodic scale models derived from the observed pitch distributions, as well as with the melodic and harmonic interval distributions. To our knowledge, our work is the first study of traditional Georgian music which treats all these aspects together.

As a side effect of our analysis, we learned that the analysis of melodic step sizes can be very misleading if one wants to derive a scale model from audio recordings. The information about the scale-degree-defining pitches rests in the central values of the pitch groups, or (in mathematical terms) in the μ_k values of the GMMs. The fact that the most frequently used melodic step size in all chants is between 170 and 180 cents (cf. Figure 20) may offer an explanation for why the results of the analysis of Tsereteli and Veshapidze (2014) let them suggest an equidistant scale model. If one randomly drew a small number of note samples from the dataset, one would very likely draw samples for which the distance between neighboring pitch groups is close to 170 to 180 cents because they are simply more likely to be drawn. The characteristic interval size around 200 cents, which we obtain because we use **all** the samples would very likely go undetected.

As to the implications of our work for the general discourse on the analysis of tuning systems of traditional Georgian vocal music, one has to keep in mind that our models represent only the practice of a single, although exceptional, singer. One should therefore not fall into the trap of the 'availability bias' (Kahnemann, Slovik, and Tversky 1982) and assume that Erkomaishvili's tuning represents 'the' Georgian tuning system. On the other hand, as Graham (2015) points out "any theory must account for [...] the tuning system heard in the 1966 Erkomaishvili recordings [...]". In this sense, we believe that Erkomaishvili's tuning model may become an important reference.

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 want to gratefully acknowledge the stimulating discussions with Malkhaz Erkvanidze, Zaal Tsereteli, and Levan Veshapidze about their work. Furthermore, we would like to thank Vlora Arifi-Müller, Stefanie Kämmerle, Moritz Berendes, Selina Sonntag, and Frank Zalkow for helping us with the annotations, the database setup, and the web-based interfaces used in this study.

References

Araqishvili, Dimitri. 2010. Svan folk songs. Chap. 3 in *Echoes from Georgia: seventeen arguments on Georgian polyphony,* edited by Rusudan Tsurtsumia and Joseph Jordania, 35–56. Nova Science Publishers, Inc.

Arom, Simha, Nathalie Fernando, and Fabrice Marandola. 2007. An Innovative Method for the Study of African Musical Scales: Cognitive and Technical Aspects. In *Smc'07, 4th sound and music computing conference,* 11–13. 11–13 July 2007. Lefkada, Greece. isbn: 978-960-6608-75-9.

Aslanishvili, Shalva. 2010. Forms of multipart singing in Georgian folk songs. Chap. 4 in *Echoes from Georgia: seventeen arguments on Georgian polyphony,* edited by Rusudan Tsurtsumia and Joseph Jordania, 57–81. Nova Science Publishers, Inc.

Burns, Edward M. 1999. Intervals, Scales, and Tuning. *The Psychology of Music:* 215–264.

Chalmers, John. 1993. *Divisions of the Tetrachord.* Hanover/New Hampshire: Frog Peak Music.

Chkhikvadze, Grigol. 2010. Main types of Georgian folk polyphony. In *Echoes from Georgia: seventeen arguments on Georgian polyphony,* edited by Rusudan Tsurtsumia and Joseph Jordania, 97–110. Nova Science Publishers, Inc.

Chokhonelidze, Evsevi. 2010. Some characteristic features of the voice coordination and harmony in Georgian multipart singing. In *Echoes from Georgia: seventeen arguments on Georgian polyphony,* 135–145. Nova Science Publishers.

Deutsch, Diana. 1970. Dislocation of tones in a musical sequence: a memory illusion. *Nature Communications* 226: 286.

— 1972. Effect of repetition of standard and comparison tones on recognition memory for pitch. *Journal of Experimental Psychology* 93: 156–162.

— 1975. The organization of short-term memory for a single acoustic attribute. In *Short-term memory,* edited by Diana Deutsch and J. A. Deutsch, 107–151. New York, NY: Academic Press.

Erkomaishvili, Anzor. 2017. *My grandfather Artem Erkomaishvili*. 356. New York: Nova Science Publishers, Inc.

Erkvanidze, Malkhaz. 2002. On Georgian scale system. In *Proceedings of the International Symposium onTraditional Polyphony,* 178–185. Tbilisi, Georgia.

— 2016. The Georgian musical system. In *Proceedings of the International Workshop on Folk Music Analysis,* 74–79. Dublin, Ireland. URL: https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1005&context=fema

Fox, Anthony. 2011. Expectation Maximization for Gaussian Mixture Distributions, Wolfram Demonstrations Project, Published: March 7 2011. URL: https://demonstrations.wolfram.com/ExpectationMaximizationForGaussianMixtureDistributions/.

Ganguli, Kaustuv Kanti, and Preeti Rao. 2019. On the perception of raga motifs by trained musicians. *The Journal of the Acoustical Society of America* 145 (4): 2418–2434. ISSN: 0001-4966.

Gelzer, Stuart. 2002. Testing a scale theory for Georgian folk music. In *Proceedings of the International Symposium onTraditional Polyphony,* 194–200. Tbilisi, Georgia.

Gogotishvili, Vladimer. 2004a. Application of authentic and plagal types of monotonic mode dominating in Georgian folk singing polyphony, 218–226. Tbilisi, Georgia.

— 2004b. On authentic and plagal types of monotonic (non-octave) scales in Georgian traditional vocal polyphony. In *Proceedings of the International Symposium onTraditional Polyphony,* 218–226. Tbilisi, Georgia.

Graham, John A. 2015. The transcription and transmission of Georgian liturgical chant. PhD, Princeton University. doi:10.1002/ejoc.201200111. arXiv: arXiv:1011.1669v3.

Heller, Eric J. 2012. Pitch Perception. In *Why you hear what you hear: an experiential approach to sound, music, and psychoacoustics,* 437–504. ISBN: 1400845580.

Hornbostel, Erich Moritz von. 1913. Melodie und Skala. *Jahrbuch der Musikbibliothek Peters für 1912:* 11–23.

Howard, David. 2007. Intonation drift in a capella soprano, alto, tenor, bass quartet singing with key modulation. *Journal of Voice* 21 (3): 300–315.

Jgharkava, Ilia. 2016. *Pearls of Georgian Chant.* CD. Produced by the Georgian Chanting Foundation & Tbilisi State Conservatoire.

Jordania, Joseph. 2010. Georgian Traditional Polyphony in comparative studies: History and Perspectives. Chap. 15 in *Echoes from Georgia: seventeen arguments on Georgian polyphony,* edited by Rusudan Tsurtsumia and Joseph Jordania, 229–248. Nova Science Publishers, Inc. http://www.novapublishers.com.

Kawai, Norie, Masako Morimoto, Manabu Honda, Eiko Onodera, and Tsutomu Oohashi. 2010. Study on sound structure of Georgian traditional polyphony. analysis of its temperament structure. In *Proceedings of the International Symposium on Traditional Polyphony,* 532–537. Tbilisi, Georgia.

Kahneman, Daniel, Slovic, Paul, & Amos Tversky. (1982). *Judgment under Uncertainty: Heuristics and Biases.* Cambridge University Press.

Mauch, Matthias, Chris Cannam, Rachel Bittner, George Fazekas, Justin Salamon, Jiajie Dai, Juan Bello, and Simon Dixon. 2015. Computer-aided melody note transcription using the Tony software: accuracy and efficiency. In *Proceedings of the international conference on technologies for music notation and representation.*

Müller, Meinard, Sebastian Rosenzweig, Jonathan Driedger, and Frank Scherbaum. 2017. Interactive fundamental frequency estimation with applications to ethnomusicological research. In *Proceedings of the AES International Conference on Semantic Audio,* 186–193. Erlangen, Germany.

Nadel, Siegfried F. 1933. *Georgische Gesänge.* 81. Lautabt., Leipzig: Harrassowitz in Komm.

Nikolsky, Aleksey. 2015. Evolution of tonal organization in music mirrors symbolic representation of perceptual reality. part-1: prehistoric. *Frontiers in Psychology* 6: 1–36.

Rosenzweig, Sebastian, Frank Scherbaum, and Meinard Müller. 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).* Delft, The Netherlands. URL: https://www.audiolabs-erlangen.de/content/05-fau/assistant/00-rosenzweig/01-publications/2019_RosenzweigSM_StableF0_ISMIR.pdf.

Rosenzweig, Sebastian, Frank Scherbaum, David Shugliashvili, Vlora Arifi-Müller, and Meinard Müller. 2020. Erkomaishvili Dataset: A Curated Corpus of Traditional Georgian Vocal Music for Computational Musicology. Transactions of the International Society for Music Information Retrieval (TISMIR), 3(1): 31–41, 2020. URL: https://www.audiolabs-erlangen.de/content/05-fau/assistant/00-rosenzweig/01-publications/2020_RosenzweigEtAl_Erkomaishvili_TISMIR_ePrint.pdf.

Scherbaum, Frank, Meinard Müller, and Sebastian Rosenzweig. 2017a. Analysis of the Tbilisi State Conservatory recordings of Artem Erkomaishvili in 1966. In *Proceedings of the International Workshop on Folk Music Analysis,* 29–36. M’alaga, Spain.

— 2017b. Rechnergestützte Musikethnologie am Beispiel historischer Aufnahmen mehrstimmiger georgischer Vokalmusik. In *Proceedings of the Jahrestagung der Gesellschaft für Informatik (GI),* 163–175. Chemnitz, Germany.

Shugliashvili, David. 2014. *Georgian church hymns, Shemokmedi School*. Georgian Chanting Foundation.

Stirzaker, David. 1999. *Probability and Random Variables, a beginner’s guide.* 368. Cambridge: Cambridge University Press.

Tsereteli, Zaal, and Levan Veshapidze. 2014. On the Georgian traditional scale, 288–295. Tbilisi, Georgia.

— 2015. Video of the presentation “The empirical research of a Georgian sound scale” (https://youtu.be/TFncneafovI). In *2015 iaml/ims congress.* New York City, USA. https://youtu.be/TFncneafovI.

Tsurtsumia, Rusudan, and Joseph Jordania. 2010. *Echoes from Georgia. Seventeen Arguments on Georgian Polyphony.* Edited by Rusudan Tsurtsumia and Joseph Jordania. New York: Nova Science Publishers, Inc. ISBN: 9781608764778.

Vavra, František, Pavel Novy, Hana Maskova, Michala Kotlikova, and Arnoštka Netralova. 2004. Morphological filtration for time series. In *Aplimat 2004,* 983–990.

Westman, Johan. 2002. On the problem of the tonality in Georgian polyphonic songs: the variability of pitch, intervals and timbre. In *Proceedings of the International Symposium onTraditional Polyphony,* 212–220. Tbilisi, Georgia.