Novel Representations of Traditional Georgian Vocal Music in Times of Online Access

Frank Scherbaum, Sebastian Rosenzweig, Reza D.D. Esfahani,
Nana Mzhavanadze, Simon Schwär, Meinard Müller

Manuscript for publication in

„Georgian Traditional Polyphony - Modern Trends and Development Perspectives“,
Ed. Rusudan Tsurtsumia and Giorgi Donadze

Version 2022/10/20

Abstract

Within the last century, the originally oral transmission of traditional (Georgian) vocal music has changed dramatically. The first mechanical and later electromagnetic field recording capabilities of sound and the associated acoustic reproducibility led to the first immense changes in the transmission mechanisms. The advent of the Internet and today's easy access to web-based audiovisual representations has enabled further significant changes in the documentation and transmission of traditional Georgian vocal music. In our paper, we explore different possibilities for novel audiovisual and partially web-based representations of traditional Georgian vocal music, making use of modern technology and computational approaches. The first example consists of a series of YouTube videos of the Tbilisi State Conservatory recordings of Artem Erkomaishvili from the year 1966. The primary goal of this example is to contribute to the preservation of this unique documentation of the performance practice and musical thinking of an outstanding singer of the last century in a modern, web-based form. With the second example, we want to introduce and demonstrate new tools to engage with recently collected multi-media (audio-, video-, and larynx microphones) field recordings of traditional Georgian vocal music. In this context, to overcome the limitations caused by the representation of non-tempered music in Western 5-line staff notation, we (semi-automatically) analyze the individual tuning system for each song and visualize the separate voices as multi-voice F0-trajectories and 'note objects' superimposed on the used F0-distribution (pitch inventory). For display, we have developed a new web-interface together with a video-audio player which combines the playback of video with an existing multitrack audio player to control the gains of the individual voices in real-time. This allows completely new ways to immerse in and interact with these types of recordings not only for documentation and analysis, but also for educational purposes. Finally, in the third example, we illustrate how these tools can be used to generate internet-based audiovisual teaching/learning scenarios in which students can sing along with selected individual voices or a virtual ensemble.

1. Introduction

Traditional Georgian polyphonic vocal music, which was originally passed down from generation to generation by oral tradition, is currently facing the challenge of the disappearance of the old ways of immersion in traditional music making. In contrast, the international interest of scientists and music lovers in traditional Georgian singing has been growing since the UNESCO declaration as part of the "Intangible Cultural Heritage of Humanity" (2001)[1)]. Professional and amateur ensembles around the world have included traditional Georgian vocal music into their repertoires, 'music tourism' to Georgia has been booming in recent years, and Georgian singing teachers regularly perform and teach abroad. Moreover, new ways of merging traditional Georgian music with various other trends of music making have emerged, e.g., in the form of Georgian folk-fusion music (Lomsadze 2019). Fortunately, one can also observe (and listen to) a new generation of talented young Georgian musicians who devote themselves to cultivating traditional Georgian music.

Therefore, it may not seem that traditional Georgian vocal music as such is an endangered genre. What has obviously changed, however, are the ways in which the music is passed on from teachers to students, especially since the time of the corona pandemic, when personal contacts had to be minimized. Nowadays, virtual internet meetings have become a popular teaching approach, and studying from scores and/or YouTube videos has become a common learning method. At a higher sophistication level, teaching CDs, often offered in sets with scores in 5-line staff notation, provide audible access to individual voice tracks (usually accompanied by a stereo mix of all voices). All of these are important tools, created with a great deal of love for the music and lots of effort, which deserve grateful recognition. But should we leave it at that? Are these tools still the most appropriate form of representing traditional Georgian music, or should we use different approaches today?

The Austrian composer Gustav Mahler (1860-1911) is (wrongly) quoted as having said "Keeping tradition alive does not mean to pray to the ashes but to pass on the flame."[2)] The question we have asked ourselves in this context is how computer-based methods and web-based techniques developed in our research project "Computational Analysis of Traditional Georgian Vocal Music (GVM)"[3)] can contribute to "passing on the flame" in ways that do justice to the essential musical aspects of the traditional music and overcome some of the limitations of current teaching materials.

With the present paper, we want to discuss three different case studies of audiovisual representations of traditional Georgian vocal music making use of modern technology and computational approaches, each of which is geared towards a different purpose. With the first example, discussed in Section 2, we want to demonstrate how computational tools and audiovisual representations can contribute to the preservation of an important set of historic recordings in a digital and more interactive form. The second example, discussed in Section 3, illustrates how a newly developed web-interface allows users to engage with a new set of multi-media (audio-, video-, and larynx microphones) field recordings of traditional Georgian vocal music (GVM dataset) in a variety of ways. In this context, we also demonstrate an intuitive visualization framework that avoids the representation of the music in western 5-line staff notation. Finally, in the third example, discussed in Section 4, we illustrate how modern digital tools can be used to generate internet-based audiovisual teaching/learning scenarios in which students can sing along with selected individual voices or a virtual ensemble.

2. A new access to the Tbilisi State Conservatory recordings of Artem Erkomaishvili (1966)

Artem Erkomaishvili (1887-1967, Fig. 1) was a key representative of traditional Georgian singing of the 20th century and one of the last master singers of Georgian liturgical chant (Erkomaishvili 2017; Graham 2015). In 1966, one year before his death, he was asked to perform all voices of a series of chants to save them for posteriority[4)]. Due to the lack of fellow singers, he had to sing all voices by himself. His performances, part of which were recently remastered (Jgharkava 2016), were recorded at the Tbilisi State Conservatory using two tape recorders through what is now called overdubbing. This recording collection is the oldest of Georgian chants and therefore of outstanding value for the Georgian cultural heritage and the understanding of the tuning principles of traditional Georgian vocal music (Graham 2015; Scherbaum et al. 2020).

Figure 1 Artem Erkomaishvili (1887-1967). Photo provided by Anzor Erkomaishvili.

The starting point for its presentation as YouTube videos is the re-release of the original audio data together with estimated fundamental frequency (F0) trajectories, beat annotations, and digital scores for each of the three voices (Rosenzweig et al. 2020). The digital scores are based on transcriptions of (Shugliashvili 2014). To enable a low-barrier access to the annotated corpus for interdisciplinary collaboration, we developed a web-based interface[5)] integrating the score-following functionality presented in (Zalkow et al. 2018). The interface allows for switching between the individual voices of the performances while synchronously highlighting the sung notes in the score-based transcription. For long-term preservation and immediate access, we additionally screen-captured the interface's output of all performances, playing back the three voices and a mix of all voices sequentially. The resulting videos are hosted on YouTube and can conveniently be accessed via a playlist[6)].

The primary goal of this contribution is to preserve (in a modern form) this unique collection, documenting the performance practice and musical thinking of an outstanding singer of the last century.

3. A new audiovisual representation of multi-media field recordings of traditional Georgian vocal music

The second case study aims to explore new perspectives on the audiovisual representation of non-tempered polyphonic vocal music. For this purpose, we make use of the specific properties of a new set of multi-media field recordings of traditional Georgian vocal music which were collected by Frank Scherbaum and Nana Mzhavanadze during a field expedition in 2016 and which we refer to as the 'GVM dataset' (Scherbaum et al. 2019). It consists of video recordings and audio recordings using conventional stereo, headset, and larynx microphones, all synchronized to a common time code. Whenever possible, one singer from each voice group was recorded with a pair of headset and larynx microphones. This way, in addition to a stereo microphone recording of the whole ensemble, separate recordings were obtained for each of the (typically three) voices of the recorded performances, one larynx microphone for computational analysis and one headset microphone to capture the natural voice of the corresponding singer in high quality (Fig. 2). Larynx microphone recordings are extremely useful for the automatic calculation of F0-trajectories (pitch trajectories) since they capture a singer's voice more or less unaffected by cross-talk from neighboring singers (Scherbaum 2016).

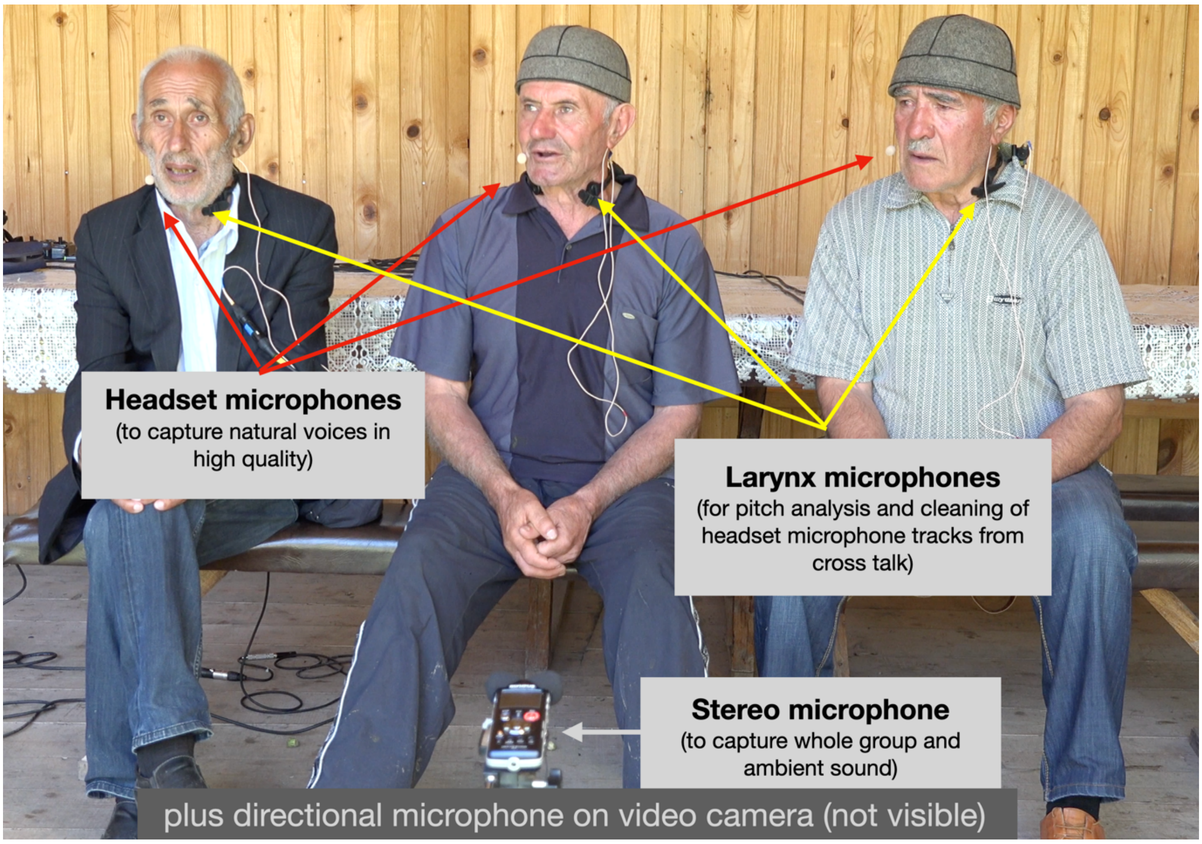
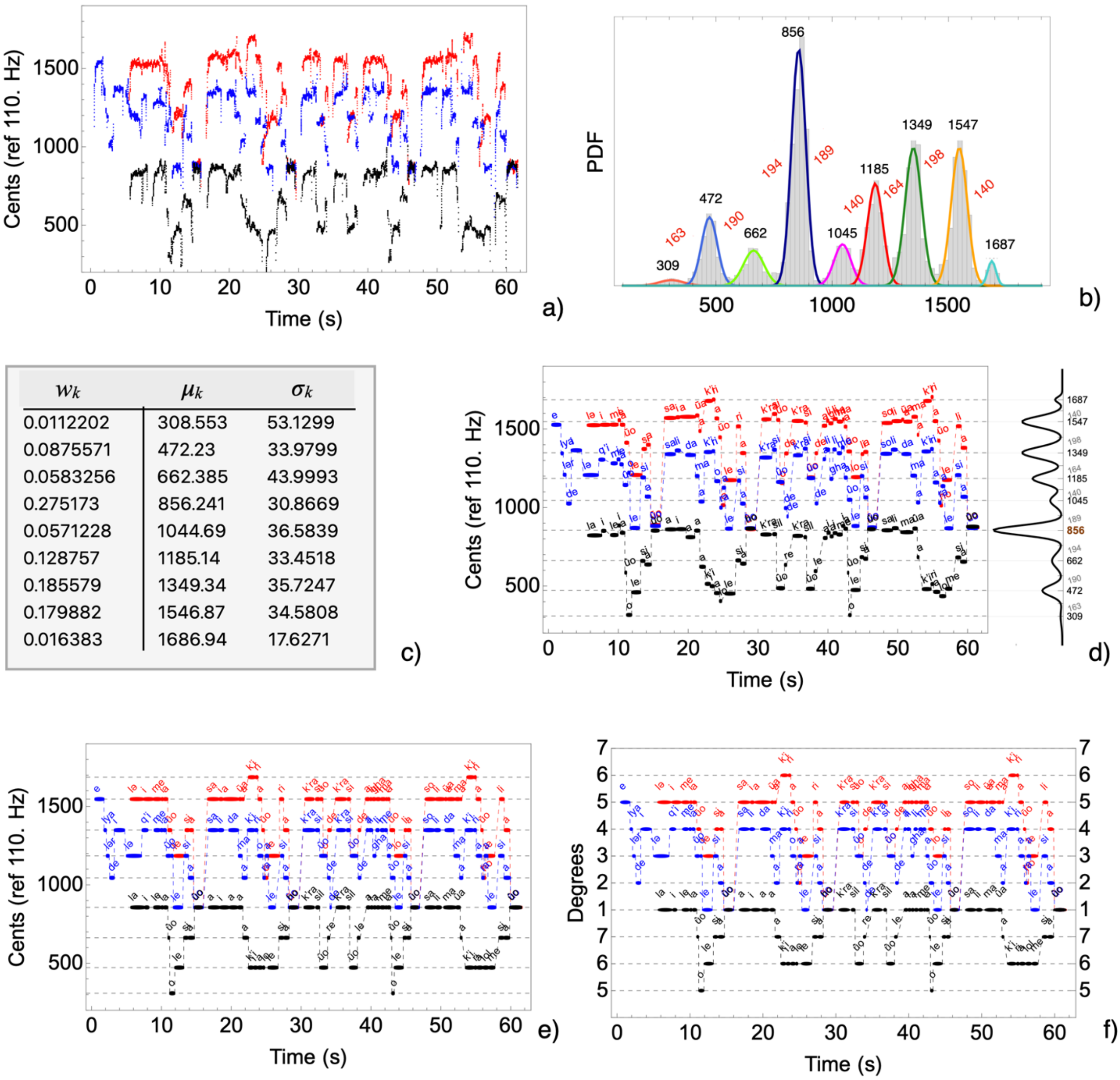


Figure 2 Typical recording setup used for the GVM collection.

3.1. Accounting for non-western tuning systems

In the field of ethnomusicology, the transcription of musical sound into a symbolic representation is commonly considered a fixed part of any analysis of oral tradition music (Ciantar 1996). Although there is agreement amongst ethnomusicologists that the tonal organization of traditional Georgian vocal music does not correspond to a 12-tone equal temperament (12-TET) tuning system (Jordania 2022), most transcriptions still use Western 5-line staff notation. Although it forces the music into an inappropriate tuning system and ignores microtonality (at least in its unmodified form). Western notation remains popular, possibly because it compresses information very effectively and can capture many aspects of the music such as dynamics and temporal structure. But does that mean it is still the best approach for all purposes?

In Fig. 3 we explore a number of alternative visual representation forms which have become possible because of the specific properties of the GVM dataset. Our goal in this context is to visualize the individual voices in a form that is intuitive to understand and which requires essentially no assumptions regarding the used tuning system. The only assumption that we make is that the cents scale (Ellis 1885) is appropriately representing how humans perceive differences in pitch (with 100 cents corresponding to 1 semitone). Since the GVM dataset contains separate larynx microphone recordings from each of the singer's voices, which are essentially unaffected by cross-talk from other voices, it is straightforward to generate individual F0-trajectories using monophonic pitch tracking algorithms such as PYIN (Mauch & Dixon 2014).



w_k	μ_k	σ_k
0.0112202	308.553	53.1299
0.0875571	472.23	33.9799
0.0583256	662.385	43.9993
0.275173	856.241	30.8669
0.0571228	1044.69	36.5839
0.128757	1185.14	33.4518
0.185579	1349.34	35.7247
0.179882	1546.87	34.5808
0.016383	1686.94	17.6271

Figure 3 a) F0-trajectories for the song "Elia Lrde'' (GVM031). b) F0-histogram and Gaussian mixture distribution (one colored Gaussian per pitch cluster). c) Parameters of the Gaussian mixture model. d) Display of the sequences of note objects produced by the Tony program (Mauch et al. 2015) (which was also used to assign the lyrics). e) Same sequence of note objects as in d) with pitches projected onto the mean value of their corresponding clusters. f) Same sequence of notes as in e) but with vertical scale changed to scale degrees.

Fig. 3a shows the raw F0-trajectories for the song ''Elia Lrde'' from Svaneti (GVM031)[7)]. The red, blue, and black curves correspond to the top, middle, and bass voice, respectively. Raw F0-trajectories include all the microtonal details of a particular recording of a song, without making any assumptions regarding the underlying tuning system.

To obtain a first visual impression of the actual melodic tuning system used, we combine all F0-samples from all three voices of the song into a single 'pitch histogram' (gray-shaded histogram in Fig. 3b) . Subsequently, we model the corresponding F0-sample distribution by a so-called Gaussian Mixture model (GMM). This is simply a weighted sum of individual Gaussian distributions $\mathcal{N}(\mu, \sigma^2)$ with mean value μ and standard deviation σ, one for each pitch group. In the case of K pitch groups, this results in a representation as $\Sigma_k\, w_k\, \mathcal{N}(\mu, \sigma^2)$. The mean values of the individual Gaussians (the μ_k) correspond to the center values of the individual pitch clusters in the pitch histogram while the standard deviations of the Gaussians (the σ_k) define the associated pitch variability. The w_k represent the individual weighting factors, in other words, how much a particular sound scale degree is present in a set of pitch trajectories. The black numbers on top of the pitch clusters in Fig. 3b show the μ_k-values of the corresponding mixture elements, while the tilted red numbers between two pitch groups correspond to the corresponding intervals in cents. The corresponding numerical parameters w_k, μ_k and σ_k are displayed in Fig. 3c.

The pitch histogram in Fig. 3b shows a strongly clustered structure with eight separate pitch clusters per octave (1200 cents). Each of the pitch clusters can be seen to represent an individual melodic sound scale degree. Therefore, the Gaussian mixture model calculated this way provides a very compact and purely data-driven representation of the tuning system used in the particular recording of the song which constitutes a natural reference frame to replace the Western 12 TET system for the visualization of the F0-trajectories[8)].

The raw F0 trajectories, however, do not seem very well suited as a replacement of western 5-line staff notation since they usually contain numerous artifacts which are not perceived as musical notes. Roughly speaking, one could say that the visual image of raw F0-trajectories is still too complex for practical purposes, such as teaching or learning a song. One way to reduce the complexity of the visualization and to approximate better what is perceived as 'note objects' is to remove sliding phases and other artifacts by applying suitable filtering techniques (Rosenzweig et al. 2019). An alternative approach, in which a sequence of note objects is generated from an F0-trajectory using a Hidden Markov Model (HMM), is provided by the Tony software (Mauch et al. 2015). Even though being more time-consuming than applying filtering techniques, although the resulting note objects have been found to be very similar to the note objects obtained by morphological filtering (Scherbaum et al. 2020), the Tony software allows a user to also interactively edit and assign lyrics to the individual note objects. This is illustrated in Fig. 3d. This type of display, which we refer to as melograph plot, still captures the microtonality of the song but reduces the information to a perceptually manageable amount. It can be seen in Fig. 3d that the pitches of the individual note objects fluctuate slightly but one can still visually identify the average pitches for each scale degree. Because of these fluctuations, it will still be difficult if not impossible to sing from such a display because it does not become clear what features are ornamental or caused by pitch instabilities of the singer and what features are intended. To address this problem, in Fig. 3e we have mapped the fluctuating note pitches for all note objects of a pitch group to the corresponding group's center value, the μ_k value. This results in a piano-roll-like display type which could be used as a "normative'' representation of the song in the sense that it could be interpreted as what the singers intended to sing. It visually reflects the underlying tuning system but not the microtonal fluctuations. Finally, for example for teaching, if one is not interested in actual pitches but only the scale degrees, one could add some additional musicological interpretation and change the vertical coordinate system to integer numbers representing scale degrees (in this case with reference to the pitch of the final bass note which is assigned scale degree 1), similar to how some singing teachers use their fingers to indicate a particular scale degree.

3.2. The GVM player

Within the GVM (Georgian Vocal Music) project[9)] and specially designed to explore new perspectives for audiovisual representation of the GVM dataset, we modified the pywebaudioplayer of (Pauwels & Sandler 2018) to what we refer to as 'GVM player'[10)]. The original player allows for individual control of the audio tracks' volumes. We extended this player by a component to synchronously play back a video stream. This opens up new paths for interacting with the music, e.g., for educational purposes. By amplifying or attenuating individual voices, various "pseudo ensemble tracks'' can be easily generated, greatly extending the famous 'Music Minus One' concept to the GVM collection. It also overcomes a limitation of currently popular educational material, such as YouTube videos and teaching CDs, which have very little if any options to interact with the music. Furthermore, by using the different visualization perspectives discussed in Section 3.1 and illustrated in Fig. 3, we also try to do justice to the non-tempered tuning systems used by many traditional Georgian singers. This is illustrated in more detail in Section 3.3.3.

3.3. Display Modes

Research and experience indicate that learning is facilitated when teachers use a variety of techniques that are purposefully selected to achieve particular learning goals (National Research Council 2002). To accommodate this strategy in the current context, we have implemented prototypes of five different display modes which can be seen as representing different scenarios. All of these can be accessed through what we refer to as the 'GVM Interface'[11]. For the purpose of the discussion of all the different display modes, we use five songs from a collection for which the raw field recordings were presented by (Scherbaum et al. 2019) and for which the original (unprocessed) recordings are easily accessible through the web-based trackswitch interface (Werner et al. 2017).[12]

The song selection starts from the top panel of GVM interface's main menu which is shown in Fig. 4 and from which different songs and display modes can be selected.

No	GVMID	Song Name	Recording Session (Video)	Audio Mix (Audio)	F0 Trajectories (Video)	Pseudo Score (Video)	Karaoke (Video)	Meta Information (PDF)
1	GVM019	Dale Koja	Play	Play	Play	Play	Play	Show Info
2	GVM009	Batonebis Nanina	Play	Play	Play	Play	Play	Show Info
3	GVM017	Chven Mshvidoba	Play	Play	Play	Play	Play	Show Info
4	GVM097	Kriste Aghsdga	Play	Play	Play	Play	Play	Show Info
5	GVM031	Elia Lrde	Play	Play	Play	Play	Play	Show Info

Teaching/Learning Scenario

No.	Identifier	Song	Play
1	MID001	Guruli Sabodisho	Click

Figure 4 Main Menu of the GVM-Interface.[11] The bottom panel corresponds to the Teaching/Learning scenario discussed in Section 4.

3.3.1. Recording Session Mode

In this display mode, the recording location, the relative position of the singers as well as their (non-verbal) communication during the recording session can be observed while listening to their performance (Fig. 5).

Figure 5 In the 'Recording Session Mode', the long shot video of the original recording session of the selected song can be watched while listening to the recording of the stereo microphone in front of the singers.

3.3.2. Audio Mix Mode

Selecting this mode starts the original pywebaudioplayer (Pauwels & Sandler 2018) (Fig. 6) for the three headset microphone tracks, for which the cross talk has been somewhat reduced by using the information about the voice activity of each singer contained in the corresponding larynx microphone recording. This mode is intended for users who are not interested in the video and want to spare themselves the overhead of the video player.

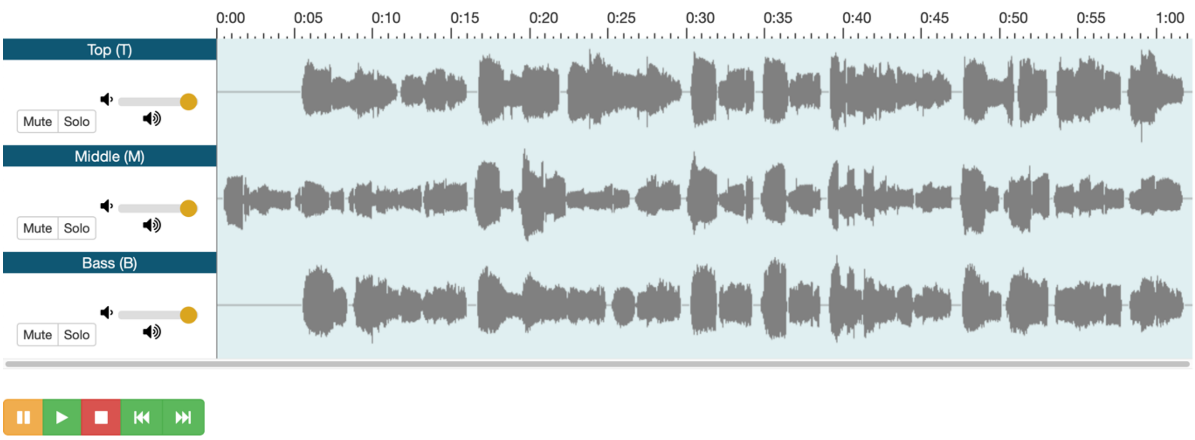


Figure 6 Interface of the pywebaudioplayer (Pauwels & Sandler, 2018).

3.3.3. F0-Trajectory Mode

In the ‘F0-trajectory Mode’, which is the most elaborate display mode implemented in the GVM player framework so far, the GVM player is started with three different display panels as shown in Fig. 7. The bottom panel corresponds to the interface of the pywebaudioplayer also shown in Fig. 6. The left top panel shows close-ups of the three singers’ faces with the lyrics of the song displayed as subtitles. To the right, the F0-trajectories (pitch tracks) are displayed together with the multi-voice note tracks for the sung notes in a “scrolling mode” where a time window of roughly 20 seconds duration (including the actual playhead-cursor position) is shown and updated at regular intervals (similar to page turning of sheet music or scores). On startup, this display panel shows the F0-trajectories of the complete song which during playback is replaced by the shorter 20 second time window. Fig. 7. shows an annotated screen image of the playback of our running example, the song “Elia Lrde”, taken at roughly 9 seconds into the song.

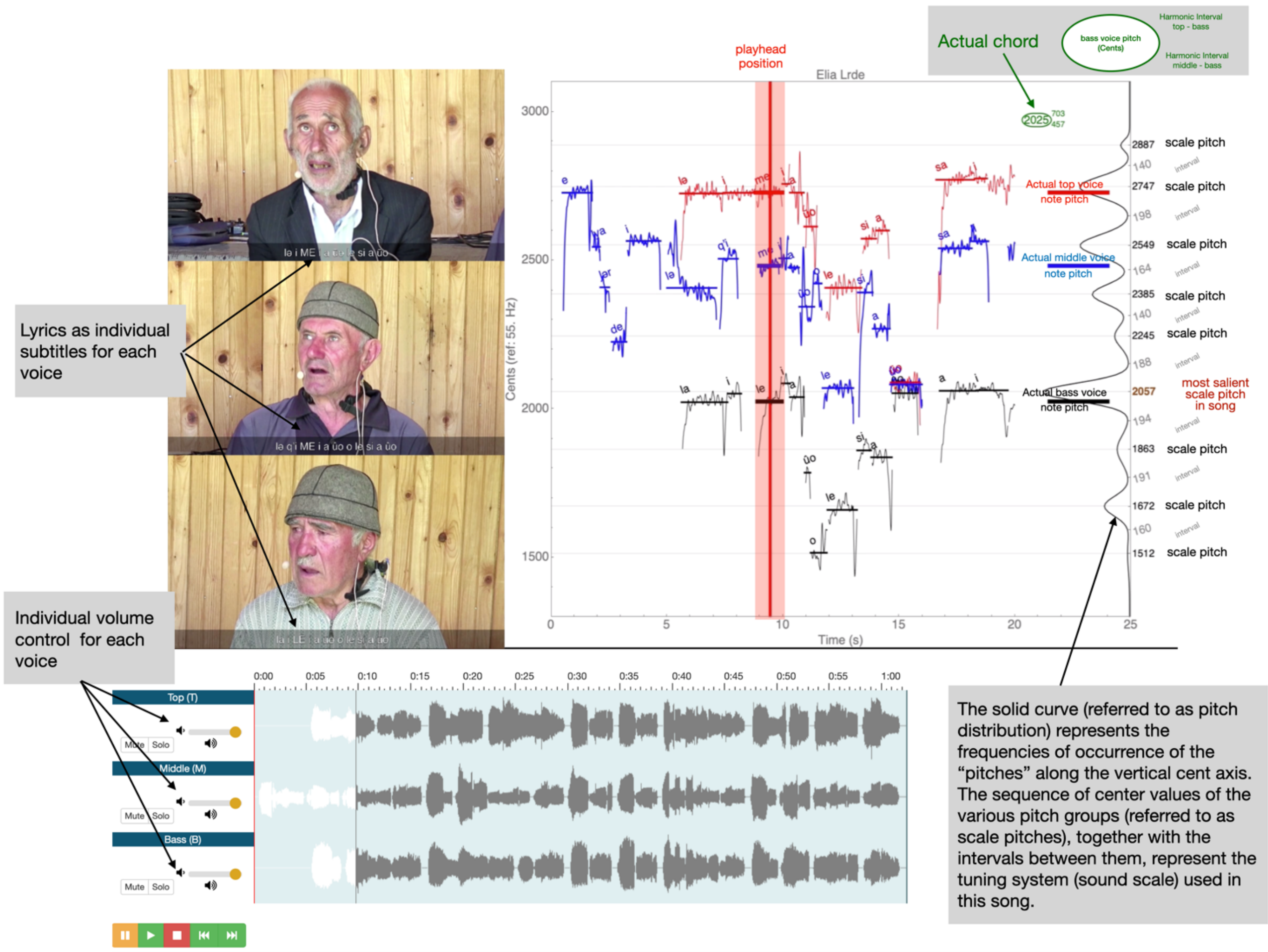


Figure 7 Screen image of the GVM player in ‘F0-trajectory display mode’, taken at roughly 9 seconds into the playback of the song “Elia Lrde”. For details see text.

In the ‘F0-trajectory display mode’, the display, in addition to showing the F0-trajectories, contains a number of analytical tools. The red vertical cursor (here for better visibility superimposed by a transparent red bar) marks the actual position of the playhead, in other words identifies what one hears at a particular moment. The solid curve on the right (referred to as pitch distribution) represents the frequencies of occurrence distribution of the “pitches”[13] along the vertical cents axis. The sequence of center values of the various pitch groups, referred to as scale pitches, correspond to the μ_k values of the Gaussian mixture model discussed in Section 3.1. Together with the intervals between them, they represent the tuning system (melodic sound scale) used in this song. The horizontal lines superimposed on the pitch distribution shows the currently sung note pitches for the top, middle and bass voice in red, blue, and black, respectively. In the upper right of the upper right panel, the actual chord is shown by a green ellipse which indicates the bass note pitch in cents with respect to a chosen reference frequency (for the conversion from Hz to cents). The sub- and superscripts show the bass-to-middle voice and bass-to-top voice intervals, respectively. Finally, the lyrics of the song are shown next to the individual notes and as subtitles in the singer's videos (with the actually sung syllable enlarged).

The vertical axis is in cents, relative to a chosen reference frequency. As already mentioned in Section 3.1, the use of the "cents scale" (100 cents correspond to a semitone) accounts for the fact that pitch perception is logarithmic. In other words, the perceived pitch difference between two pitches being 100 cents apart is always a semitone, independent of where on the pitch scale it takes place[14)]. The sung notes (referred to as 'note objects' above) are shown by horizontal lines, color coded in red, blue, and black for the top, middle, and bass voice, respectively. The pitch values of these lines corresponds to the corresponding note pitches (as determined by the Tony program) at the time of the cursor position.

The solid black curve on the right side of the display window (marked as pitch distribution and oriented perpendicular to the vertical cents axis) describes the used pitch inventory (i.e. the distribution of F0-values) of the complete song. The area under the Gaussian bell-shaped curve for a single pitch group expresses the salience of the corresponding pitch in the song. The values of the scale pitches in cents are indicated to the right of the location of the corresponding peak of the pitch distribution function, with the scale pitch of the most salient pitch group being shown in red. The size of the intervals between two neighboring scale pitches is indicated by the tilted numbers between two pitch groups. The sequence of scale pitches is also marked in the display window by thin horizontal lines. Together with the intervals between them they represent the tuning system (sound scale) used in the displayed song in a purely "descriptive'' sense, i.e. without making any prior assumptions about the tuning. This approach is in line with the "philosophy" to try to only show what is in the observed data and to leave the interpretation of these measurements and further analysis to the user.

The actually sung pitch values of the individual voices that a listener hears at a particular instance of time are displayed in real-time by color-coded horizontal bars superimposed on the pitch distribution function. This, in addition to what the listener hears at that moment, provides real-time information if a particular voice pitch is below, at, or above a scale pitch.

There is one additional aspect of Fig. 7 which needs to be mentioned. It relates to the harmonic content of a song and the fact that a traditional Georgian song is more than the combination of three melodies sung at the same time. Teaching CDs usually provide the individual voices as isolated melodies. Details of the harmonic interaction between singers, which in live performances can be perceived as very fine-grained mutual (harmonic) intonation adjustments, usually get lost this way. To address this aspect in the GVM player, the actually sung chord (bass voice scale degree and harmonic intervals between bass and middle and bass and top voice in cents in subscript and superscript, respectively) are displayed in real-time by the green chord symbol in the upper right of the display panel. This information not only helps to train harmonic perception („vertical thinking") but can also be useful for real-time harmonic analysis of a song.

The fact that the GVM player can simultaneously play the audio and the videos of the singers has the side effect that it allows to engage with the recorded material of a performance not only acoustically but also visually and to observe the non-verbal interaction between singers. To facilitate the perception of a song even further, we display the lyrics of a song for each voice not only on top of individual 'note objects', but also as subtitles in the videos of the individual singers. This can be done in Georgian letters as well as in transcribed Roman letters (as it is done here).

3.3.4. Pseudo-Score Mode

In this display mode, the note pitches have all been mapped to the center pitches of their corresponding pitch group, and the display of the F0-trajectories is suppressed (as an unnecessary detail in this setting) (Fig. 8). In contrast to the F0-trajectory mode which displays exactly the pitches sung by the singers, this mode is an interpretation of what pitch the singers might have wanted to sing, using the pitch group means as a reference. The motivation for this mode comes from the observation that pitch perception is categorical (c.f. Ganguli & Rao 2019). It is a mode that makes it easier for students to recognize which scale degree was sung. This is why we refer to it tentatively as 'Pseudo-Score mode'. We anticipate this to be the most attractive mode to learn a song, in particular in combination with the volume control for the individual audio tracks.

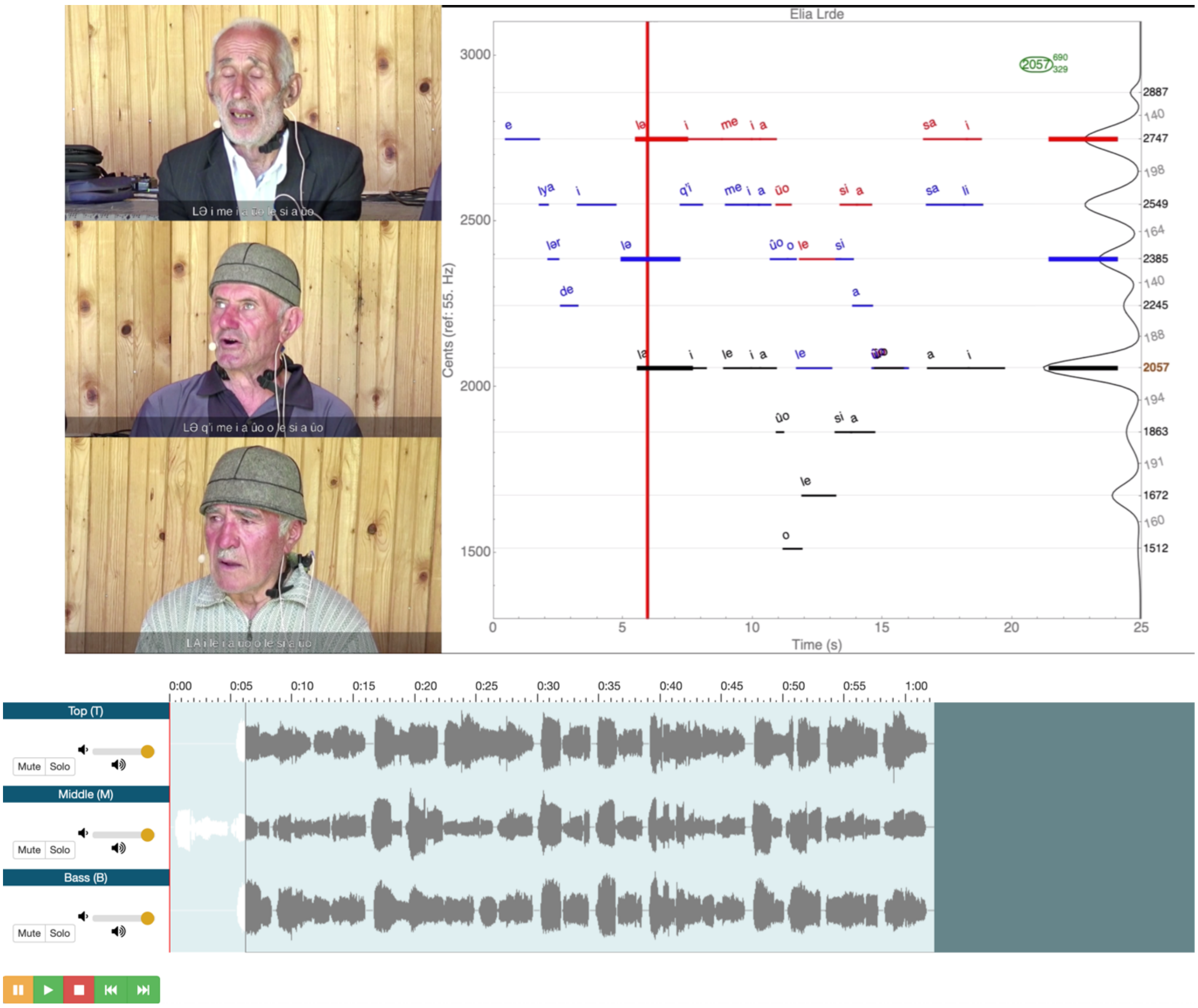


Figure 8 Screen image of the 'Pseudo-Score Mode'.

3.3.5. Karaoke Mode

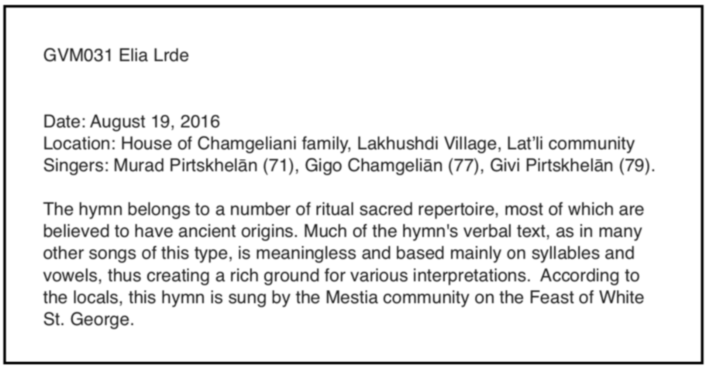
In this mode, we mimic a Karaoke situation by displaying only the lyrics together with close-ups of the singers' faces (Fig. 9). Like in the F0-trajectory mode and the Pseudo-Score mode, the volume of the audio tracks can be individually adjusted (and not just muted as e.g. in the trackswitch interface).

Figure 9 Screen image illustrating the 'Karaoke mode'.

3.3.6. Meta Information

Each recording session during the 2016 field expedition was accompanied by extensive interviews not only with the singers but also with other informants from the villages to collect contextual information regarding the background and history of the singers, about local customs, etc.

Selecting the ‘Show Info button’ under the Meta Information entry will bring up a small subset of this information, e. g., the singers’ names and recording dates (Fig. 10). The complete field reports can be obtained from the LaZAR-Database.[15)]


GVM031 Elia Lrde

Date: August 19, 2016
Location: House of Chamgeliani family, Lakhushdi Village, Lat’li community
Singers: Murad Pirtskhelān (71), Gigo Chamgeliān (77), Givi Pirtskhelān (79).

The hymn belongs to a number of ritual sacred repertoire, most of which are believed to have ancient origins. Much of the hymn's verbal text, as in many other songs of this type, is meaningless and based mainly on syllables and vowels, thus creating a rich ground for various interpretations. According to the locals, this hymn is sung by the Mestia community on the Feast of White St. George.

Figure 10 Meta information window for the song “Elia Lrde”.

4. Teaching/Learning Scenario

Although originally designed for accessing the GVM dataset (Scherbaum et al. 2019), the GVM Interface has also turned out to be very useful for the generation of powerful internet-based teaching/learning scenarios. During the lockdown phase of the Corona pandemic, Nana Mzhavanadze, who, besides being an ethnomusicologist, is a professional singer and experienced teacher of traditional Georgian vocal music, was forced (like many of her colleagues) to abandon her traditional ways of face-to-face teaching. Inspired by the 1966 recordings of Artem Erkomaishvili's State Conservatory, we tried to cope with this situation by extending the overdubbing technique used in 1966 from audio to audiovisual recordings and to exploit the capabilities of the GVM player. Instead of several tape recorders as in 1966, several instances of the freely available QuickTime video player were used. After Nana had recorded the first voice through her computer’s microphone and video camera, the recording was played back to her over headphones, using one instance of the video player. She then recorded the second voice (video and audio) with a second instance of the player. This procedure was then repeated for the third voice. The three videos were then synchronized to a common time code and cut to a common length. Subsequently, the audio tracks were analyzed in the same way as the recordings in the GVM dataset, in this case by using the Tony software to produce pitch and note tracks and to annotate the lyrics. From the annotated note tracks, the subtitle files were calculated and subsequently added to the videos. The resulting video and audio tracks were finally made accessible through the GVM player. A screen image for an example of this scenario is shown in Fig. 11.

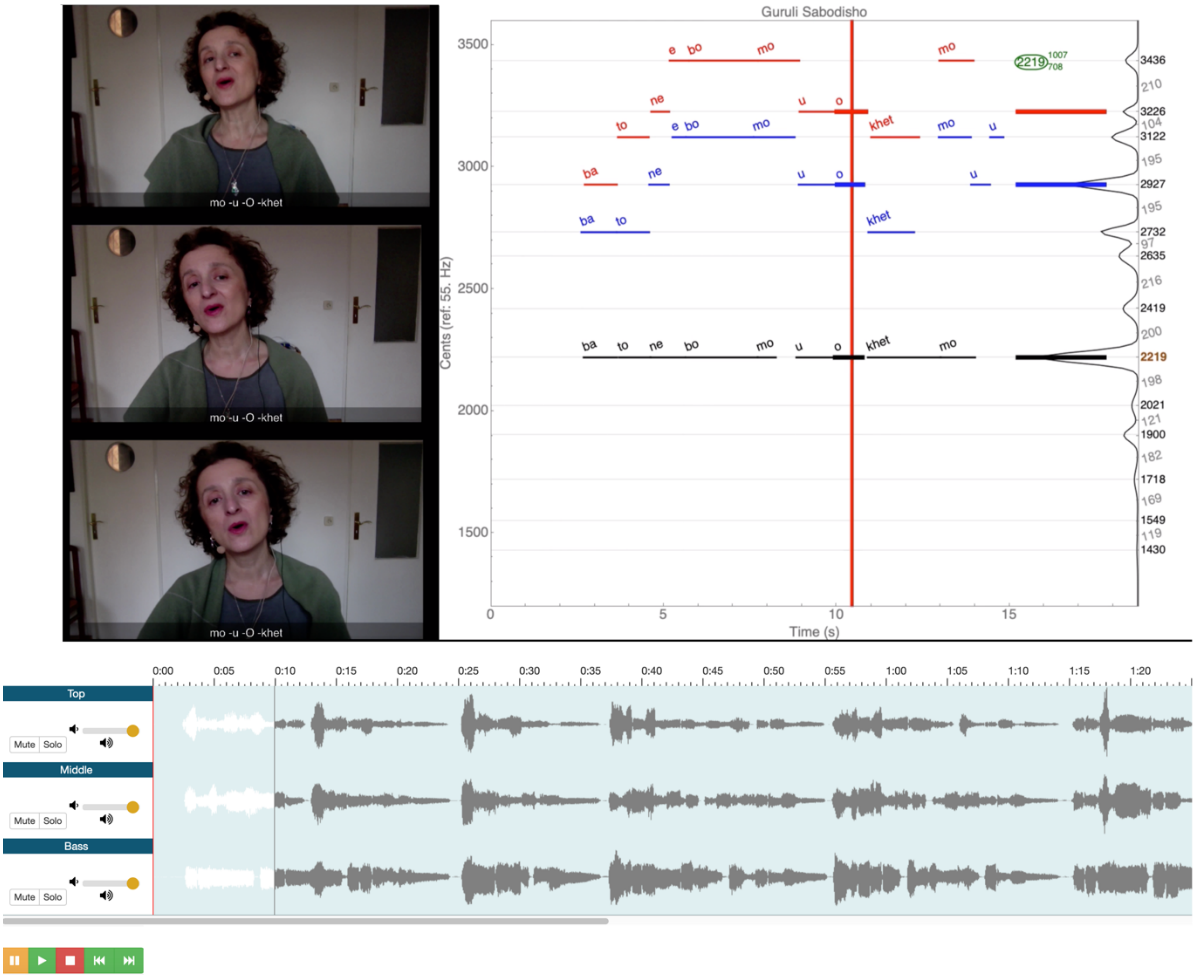


Figure 11 Nana Mzhavanadze singing a single-singer-version of Guruli Sabodisho as a potential teaching scenario.[16)]

As a result, students were able to sing along to selected individual voices and/or virtual ensembles for which the volumes of the individual voices could be interactively controlled in real-time while simultaneously watching the videos of the teacher and reading the lyrics of the songs as subtitles.

5. Discussion

In the current study we have taken a computational perspective on ethnomusicology to explore new ways of representing traditional Georgian vocal music for preservation, documentation, analysis, and educational purposes. The goal of this exercise was not to come up with finished products but to explore the potential of modern state-of-the-art computational tools to process, analyze and present a recent collection of multi-media field recordings, while identifying existing technical and conceptual challenges. We believe that the presented examples demonstrate for once that the advent of the internet and today's easy access to web-based audiovisual representations and computational tools can have enormous benefits for the analysis, documentation, and preservation of non-western oral music traditions, in particular when ethnomusicological field expeditions aim to extend their recording setup to multi-media and multi-channel recordings which would include conventional audio, video, and larynx microphones (for analysis purposes). The recent study by (Scherbaum & Müller 2022)also illustrates the benefit for studying the interaction between singers using additional heartbeat recordings.

We also believe that computational tools have matured enough to help overcome the problems related to the representation of non-tempered traditional Georgian vocal music by simply using static Western 5-line staff notation. This notation is based on the assumption of a 12 tone-equal-temperament (12-TET) tuning system which is inappropriate to represent the non-tempered tuning systems used by many traditional Georgian singers. The way we have tried to address this problem (see e.g. in Fig. 3, 4 and 8) should be seen only as a first step and a proof-of-concept of what may be done. Finally, it needs to be emphasized, that in the present paper we completely omitted the discussion of the underlying preprocessing workflow. Computational analysis of modern multimedia multi-channel field recordings still faces considerable procedural challenges in terms of being very labor-intensive. For that reason, we have restricted the number of analyzed songs in the present study to only five. Some of the processing issues are addressed in (Rosenzweig et al. 2022; Scherbaum et al. 2022). One of the most time-consuming open issues is currently the determination of the song lyrics and the association with note objects. In summary, the design goal for the GVM player was to allow users to immerse as deeply as possible into the rich polyphonic (non-tempered) soundscape of traditional Georgian singing, using state-of-the-art computer tools.

6. Acknowledgments

The authors thank Vlora Arifi-Müller and Kathrin Hedrich for their assistance in preparing the data, annotations, and videos for the Erkomaishvili corpus. This work was supported by the German Research Foundation (DFG MU 2686/13-1, SCHE 280/20-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg and the Fraunhofer Institut für Integrierte Schaltungen IIS.

7. Notes

1) https://ich.unesco.org/en/RL/georgian-polyphonic-singing-00008 [accessed, 20 October 2022]

2) According to the blogpost https://falschzitate.blogspot.com/2017/06/tradition-ist-die-weitergabe-des-feuers_10.html [accessed, 20 October 2022], true author of this metaphor is the French socialist Jean Jaurè, who is said to have coined it in a speech to the French Parliament on 21 January 1910.

3) Funded by the German Research Foundation (DFG MU 2686/13-1, SCHE 280/20-1) since 2018.

4) Since chanting was prohibited during the Soviet era, these chants might have been lost without these recordings.

5) https://www.audiolabs-erlangen.de/resources/MIR/2019-GeorgianMusic-Erkomaishvili [accessed, 20 October 2022]

6) https://www.youtube.com/playlist?list=PLs9CCVIpRRzi2cw_jJ2NH4vH7AZ1k-Mzv [accessed, 20 October 2022]

7) Recorded on August 19, 2016 at the House of Chamgeliani family, Lakhushdi Village, Lat'li community. Singers: Murad Pirtskhelān (71) middle voice, Gigo Chamgeliān (77) top voice, Givi Pirtskhelān (79) bass voice. The hymn belongs to a number of ritual sacred repertoire, most of which are believed to have ancient origins. Much of the hymn's verbal text, as in many other songs of this type, is meaningless and based mainly on syllables and vowels, thus creating a rich ground for various interpretations. According to the locals, this hymn is sung by the Mestia community on the Feast of White St. George.

8) It is worth noting that the raw F0 trajectories still contain sliding phases and other (non-musical artifacts) , and can also be subject to continuous slow pitch drifts. Therefore, an in-depth analysis of the tonal organization of a set of recordings, usually requires a more elaborated workflow (Scherbaum et al., 2022; Scherbaum & Mzhavanadze, 2020).

9) https://www.audiolabs-erlangen.de/fau/professor/mueller/projects/gvm
[accessed, 20 October 2022]

10) https://www.audiolabs-erlangen.de/resources/MIR/GVMPlayerInfo [accessed, 20 October 2022]

11) https://www.audiolabs-erlangen.de/resources/MIR/GVMPlayer/ [accessed, 20 October 2022]

12) https://www.audiolabs-erlangen.de/resources/MIR/2018-ISMIR-LBD-ThroatMics
[accessed, 20 October 2022]

13) Since pitch is a psychoacoustic quantity which can not be measured directly, F0-trajectories can be seen as a quantitative approximation to the actually sung pitches. It needs to be emphasized again, however, that F0-trajectories not only capture sung notes but also other details of the intonation process, as well as other acoustic events such as swallowing and clearing the throat. For example, the screen image shows sliding phases at the beginning and ends of notes, which are quite typical for the Svan intonation style.

14) This is already a considerable simplification in relation to a western score notation where the perceived difference between two notes also depends on the key of the score.

15) https://lazardb.gbv.de/search
[accessed, 20 October 2022]

16) https://www.audiolabs-erlangen.de/resources/MIR/GVMPlayer/MID_0.html
[accessed, 20 October 2022]

8. References

Ciantar, Philip. 1996. “Styles of Transcription in Ethnomusicology”. MA thesis, Durham University, Dept. of Music. http://etheses.dur.ac.uk/2024/1/2024_31.pdf [accessed, 18 August 2022]

Ellis, Alexander J. 1885. “On the Musical Scales of Various Nations”. *Journal of the Society of Arts*, xxxiii, 486 - 527. https://de.scribd.com/document/239042170/Alexander-J-Ellis-On-the-Musical-Scales-of-Various-Nations-Journal-of-the-Society-of-Arts-for-March-27-1885-vol-xxxiii [accessed, 18 August 2022]

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

Ganguli, Kaustuv Kanti & Preeti Rao. 2019. “On the perception of raga motifs by trained musicians”. *The Journal of the Acoustical Society of America*, 145(4), 2418–2434. https://doi.org/10.1121/1.5097588 [accessed, 18 August 2022]

Graham, John A. 2015. “The transcription and transmission of Georgian Liturgical chant”. Ph.D. thesis, Princeton University, Princeton. https://dataspace.princeton.edu/handle/88435/dsp01zk51vk16s [accessed, 18 August 2022]

Jgharkava, Ilya. 2016. “Pearls of Georgian Chant: Artem Erkomaishvili”. Georgian Chanting Foundation & Tbilisi State Conservatory.

Jordania, Joseph. 2022. “Continuing Discussions on Scale Systems in Georgian Traditional Music”. In: *Anzor Erkomaishvili and Contemporary Trends in the Performance and study of Georgian traditional and sacred music*. Edited by: J. Jordania & R. Tsurtsumia . Cambridge Scholars Publishing.

Lomsadze, Teona. 2019. “When tradition meets modernity–Georgian folk-fusion music”. *Folk Life*, 57(2), 122–140. https://doi.org/10.1080/04308778.2019.1656791 [accessed, 18 August 2022]

Mauch, Matthias, Chris Cannam, Rachel Bittner, George Fazekas, Justin Salamon, Jiajie Dai, Juan Bello & Simon Dixon. 2015. *Computer-aided Melody Note Transcription Using the Tony Software : Accuracy and Efficiency*. Proceedings of the First International Conference on Technologies for Music Notation and Representation, 8. https://code.soundsoftware.ac.uk/projects/tony/

Mauch, Matthias & Simon Dixon. 2014. *PYIN : A fundamental frequency estimator using probabilistic threshold distributions*. IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2014, 659–663.

National Research Council. 2002. “Learning and Understanding: Improving Advanced Study of Mathematics and Science in U.S. High Schools”. In: *Learning and Understanding.* Washington, DC: The National Academies Press. https://doi.org/10.17226/10129 [accessed, 18 August 2022]

Pauwels, Johan, & Mark B. Sandler. 2018. *pywebaudioplayer : Bridging the gap between audio processing code in Python and attractive visualizations based on web technology*. Web Audio Conference WAC-2018, September 19–21, 2018, Berlin, Germany.

Rosenzweig, Sebastian, Frank Scherbaum & Meinard Müller. 2019. *Detecting Stable Regions in Frequency Trajectories for Tonal Analysis of Traditional Georgian Vocal Music*. 20th International Society for Music Information Retrieval Conference, Delft, The Netherlands, 2019., 352–359.

Rosenzweig, Sebastian, Frank Scherbaum & Meinard Müller. 2022. “Computer-Assisted Analysis of Field Recordings: A Case Study of Georgian Funeral Songs”. *ACM Journal on Computing and Cultural Heritage*, in press, https://dl.acm.org/doi/abs/10.1145/3551645 [accessed, 18 August 2022]

Rosenzweig, Sebastian, Frank Scherbaum, David Shugliashvili, Vlora Arifi-Müller & 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*, 3(1), 31–41. https://doi.org/10.5334/tismir.44 [accessed, 18 August 2022]

Scherbaum, Frank. 2016. *On the benefit of larynx-microphone field recordings for the documentation and analysis of polyphonic vocal music*. Proc. of the 6th International Workshop Folk Music Analysis,15 - 17 June, Dublin/Ireland, 80–87. https://arrow.tudublin.ie/cgi/viewcontent.cgi?article=1010&context=fema [accessed, 18 August 2022]

Scherbaum, Frank & Meinard Müller. 2022. “From Intonation Adjustments to Synchronisation of Heart Beat Variability: Singer Interaction in Traditional Georgian Vocal Music”. *Musicologist*, submitted.

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

Scherbaum, Frank, Nana Mzhavanadze, Simha Arom, Sebastian Rosenzweig & Meinard Müller. 2020. “Tonal Organization of the Erkomaishvili Dataset: Pitches, Scales, Melodies and Harmonies“. In: *Computational Analysis Of Traditional Georgian Vocal Music* (Issue 1). Edited by: Scherbaum, F. https://publishup.uni-potsdam.de/frontdoor/index/index/docId/47614 [accessed, 18 August 2022]

Scherbaum, Frank, Nana Mzhavanadze, Sebastian Rosenzweig & Meinard Müller. 2019. *Multi-media recordings of traditional Georgian vocal music for computational analysis*. 9th International Workshop on Folk Music Analysis, 2-4 July 2019. https://www.uni-potsdam.de/fileadmin/projects/soundscapelab/PapersMusic/2019/Scherbaum_etal_FMA2019_final.pdf

Scherbaum, Frank, Nana Mzhavanadze, Sebastian Rosenzweig & Meinard Müller. 2022. ”Tuning Systems of Traditional Georgian Singing Determined From a New Corpus of Field Recordings”. *Musicologist*, submitted.

Scherbaum, Frank, Sebastian Rosenzweig, Meinard Müller, Daniel Vollmer & Nana Mzhavanadze. 2018. *Throat Microphones for Vocal Music Analysis*. Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR). https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/03-publications/2018_ScherbaumRMVM_GVM_ISMIR-LBD.pdf [accessed, 18 August 2022]

Shugliashvili, David. 2014. *Georgian Church Hymns, Shemokmedi School*. Georgian Chanting Foundation & Tbilisi State Conservatory.

Werner, Nils, Stefan Balke, Fabian-Robert Stöter, Meinard Müller & Bernd Edler. 2017. *trackswitch.js: A Versatile Web-Based Audio Player for Presenting Scientific Results*. Web Audio Conference WAC-2017, August 21–23, 2017, London, UK. https://www.audiolabs-erlangen.de/content/05-fau/professor/00-mueller/03-publications/2017_WernerBSME_TrackswitchJSPlayer_WAC.pdf [accessed, 18 August 2022]

Zalkow, Frank, Sebastian Rosenzweig, Johannes Graulich, Lukas Dietz, El Mehdi Lemnaouar & Meinard Müller, M. 2018. A Web-Based Interface for Score Following and Track Switching in Choral Music. Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR). https://www.audiolabs-erlangen.de/content/resources/MIR/00-2018-ISMIR-LBD-Carus/2018_ZalkowRGDMM_Carus_ISMIR-LBD.pdf [accessed, 18 August 2022]