Computer-assisted Analysis of Field Recordings: A Case Study of Georgian Funeral Songs

SEBASTIAN ROSENZWEIG, International Audio Laboratories Erlangen
FRANK SCHERBAUM, University of Potsdam
MEINARD MÜLLER, International Audio Laboratories Erlangen

Three-voiced funeral songs from Svaneti in North-West Georgia (also referred to as Zär) are believed to represent one of Georgia's oldest preserved forms of collective music-making. Throughout a Zär performance, the singers often jointly and intentionally drift upwards in pitch. Furthermore, the singers tend to use pitch slides at the beginning and end of sung notes. Musicological studies on tonal analysis or transcription have to account for such musical peculiarities, e.g., by compensating for pitch drifts or identifying stable note events (located between pitch slides). These tasks typically require labor-intensive annotation processes with manual corrections executed by experts with domain knowledge. For instance, in the context of a previous musicological study on pitch inventories (or pitch-class histograms) of Zär performances, ethnomusicologists tediously annotated fundamental frequency (F0) trajectories, stable note events, and pitch drifts for a set of 11 multitrack field recordings. In this article, we study how musicological studies on field recordings can benefit from interactive computational tools that support such annotation processes. As one contribution of this article, we compile a dataset from the previously annotated audio material, which we release under an open-source license for research purposes. As a second contribution, we introduce two computational tools for removing pitch slides and compensating pitch drifts in performances. Our tools were developed in close collaboration with ethnomusicologists and allow for incorporating domain knowledge (e.g., on singing styles or musically relevant harmonic intervals) in the different processing steps. In a case study using our Zär dataset, we evaluate our tools by reproducing the pitch inventories from the original musicological study and subsequently discuss the potential of computer-assisted approaches for interdisciplinary research.

CCS Concepts: • **Human-centered computing** → *Interactive systems and tools*; • **Applied computing** → **Sound and music computing;**

Additional Key Words and Phrases: Interactive tools, Georgia, vocal music, Zär, tonal analysis, pitch slides, pitch drift

ACM Reference format:
Sebastian Rosenzweig, Frank Scherbaum, and Meinard Müller. 2022. Computer-assisted Analysis of Field Recordings: A Case Study of Georgian Funeral Songs. *J. Comput. Cult. Herit.* 16, 1, Article 13 (December 2022), 16 pages.
https://doi.org/10.1145/3551645

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.

Authors' addresses: S. Rosenzweig and M. Müller, International Audio Laboratories Erlangen, Am Wolfsmantel 33, Erlangen, Bavaria 91058, Germany; emails: {sebastian.rosenzweig, meinard.mueller}@audiolabs-erlangen.de; F. Scherbaum, University of Potsdam, Potsdam 14469, Germany; email: fs@geo.uni-potsdam.de.

1556-4673/2022/12-ART13
https://doi.org/10.1145/3551645

1 INTRODUCTION

Georgia is home to centuries-old music traditions. In particular, Georgian polyphonic singing is listed as "Intangible Cultural Heritage of Humanity" by the UNESCO.[1] As part of an interdisciplinary and publicly funded research project,[2] computer scientists and ethnomusicologists collaborate with the goal to advance research on Georgia's musical treasure using computer-assisted approaches. In particular, one goal is to obtain a better understanding of the controversially discussed tonal organization of traditional **Georgian vocal music** (**GVM**) [6, 35, 37, 46].

In this context, three-voiced funeral songs (or dirges, also referred to as Zär) from the region Svaneti in North-West Georgia have gained special attention among ethnomusicologists since they represent one of the oldest forms of collective music-making in Georgia [11]. Zär performances exhibit two musical peculiarities, which can be observed when looking at the fundamental frequency (F0) trajectories of the singers' voices as depicted in Figure 1(a). First, the singers tend to use pitch slides at the beginning and end of sung notes (also referred to as portamento). Second, throughout a Zär performance, the singers may jointly and intentionally drift upwards in pitch by even more than 500 cents [39]. The presence of pitch slides and pitch drifts constitutes a challenge for tonal analysis or transcription, as we will illustrate in the following. An important part of the tonal analysis is the determination of pitch inventories (or pitch-class histograms) [8, 14, 16, 36, 48], which can be computed by accumulating the F0-values over time. As one can see in Figure 1(b), pitch slides and drifts may result in noisy and blurry pitch inventories, which are hard to interpret or even meaningless for tonal analysis.

To tackle this problem, one strategy is to remove pitch slides and compensate for pitch drifts prior to computing pitch inventories. Such tasks typically need to be conducted by experts with domain knowledge. In the context of four ethnomusicological studies on a set of 11 multitrack recordings of Zär performances [24, 25, 39, 40], domain experts annotated stable note events (F0-values between pitch slides) for all voices, as depicted in Figure 1(c). Subsequently, the ethnomusicologists selected note events that best reflect the pitch drift of the performances (see the black boxes in Figure 1(c)) and determined pitch drifts through polynomial curve fitting (see the black line in Figure 1(c)). After drift-correction with the (suitably normalized) drift curve, one obtains the drift-corrected stable note events as depicted in Figure 1(d) and the pitch inventory as depicted in Figure 1(e). As one can see, in contrast to the uncorrected pitch inventory from Figure 1(b), the pitch inventory based on the annotated material exhibits a sharper distribution. Through comparison of the pitch inventories for all 11 performances (determined in the same way), ethnomusicologists could show that the melodic step sizes in Zär vary between approximately 150 and 180 cents, which is an important cue for understanding the traditional Georgian tuning system [39, 41]. However, conducting such annotation processes using existing semi-automatic annotation tools is labor-intensive and requires manual corrections. This also makes it hard to conduct similar studies on larger corpora.

In this article, we show that computational tools can support the analysis of field recordings by automizing some of the labor-intensive annotation tasks under the guidance of a domain expert. As one contribution, we compiled a dataset including the multitrack recordings and the carefully crafted annotations from the musicological studies, which we release under an open-source license for research purposes.[3] As our main technical contribution to this article, we present two computational tools with visual feedback mechanisms that allow for incorporating musical expert knowledge into the different processing steps. Our first tool, based on an existing method for extracting stable regions from F0-trajectories [33], enables the user to remove pitch slides and other undesired frequency fluctuations. The method's parameters can be tuned according to musical characteristics such as the singing style. Our second tool is based on a filtering technique for musically relevant harmonic intervals (such as the unison or the fifth in GVM [4, 35, 38]) to compensate for the pitch drift of a performance. In

[1] https://ich.unesco.org/en/RL/georgian-polyphonic-singing-00008.
[2] https://www.uni-potsdam.de/de/soundscapelab/computational-ethnomusicology/active-projects/the-gvm-project.
[3] https://www.audiolabs-erlangen.de/resources/MIR/2022-GeorgianMusic-Zaer.

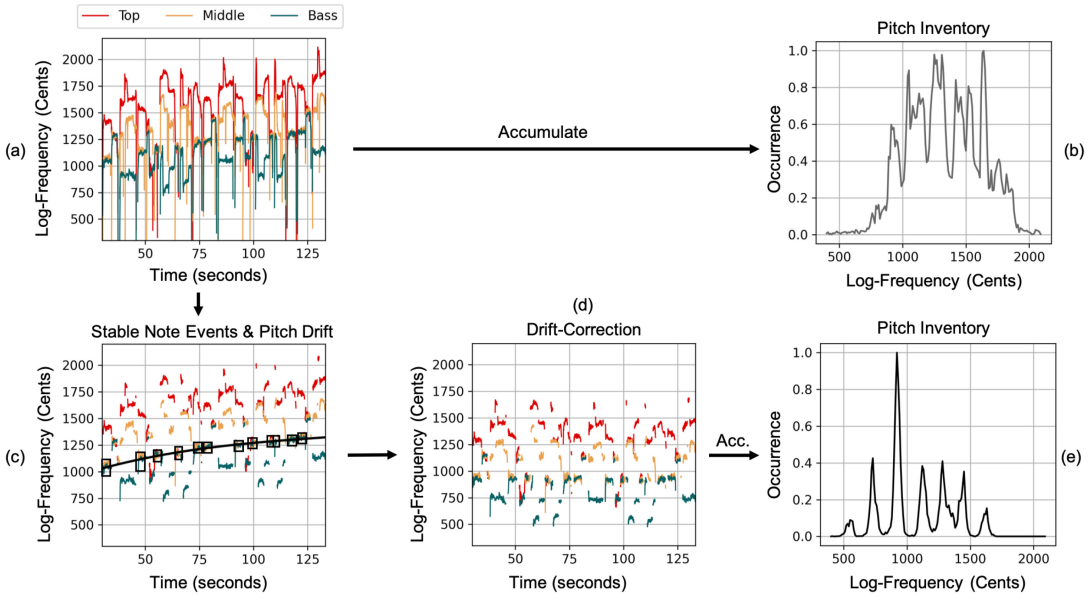

Fig. 1. Pitch inventory computation for a three-voiced Zär performance. **(a)** Original F0-trajectories. **(b)** Pitch Inventory based on (a). **(c)** Annotated stable note events and pitch drift curve (black line). **(d)** Drift-corrected stable note events. **(e)** Pitch inventory based on (d).

a case study based on our Zär dataset, we compare pitch inventories computed with our computational tools to the ones obtained from the expert annotations [24, 25, 39].

The remainder of this article is structured as follows. We describe related work in Section 2 and introduce our Zär dataset in Section 3. In Section 4, we formalize our computational tools. In Section 5, we describe our case study and discuss the potential of interactive computational tools for ethnomusicological research. Finally, we summarize our work and outline future research in Section 6.

2 RELATED WORK

Pitch slides and pitch drifts are a frequently observed phenomenon in a cappella singing, not least due to the great versatility of the human voice [45]. However, the musicological perspective on these phenomena often depends on the cultural context. For instance, pitch slides are often considered as a sign of insufficient voice control in Western amateur choral singing while being a frequently and consciously used stylistic element in other music cultures such as traditional GVM [35] or Indian Raga music [7, 13]. Similarly, pitch drifts are typically seen as unintended artifacts of tuning in "Western" ensemble singing [1, 10] while they are known to be a part of the performance practice in several "non-Western" music traditions [2, 12, 20, 22], including Georgian Zär [24, 25, 39, 40]. Thus, computational analysis of field recordings requires tools that can be adapted to the musical scenario by including musical or culture-specific knowledge [9]. Additionally, such tools need to offer suitable feedback mechanisms, e.g., visualizations or sonifications [49], which help to understand and guide computational methods.

Over the last years, a variety of tools for annotating and analyzing music recordings with a focus on "Western" music have been released [3, 17, 21, 23, 27, 30]. One of the most popular tools for transcribing monophonic audio recordings is the open-source software Tony [18]. After loading an audio file, the tool automatically computes an F0-trajectory using the algorithm pYIN [19]. Via an interactive **graphical user interface** (**GUI**) with

audiovisual feedback, the user can remove F0-values or choose alternative estimates. For transcription purposes, Tony automatically detects sung notes by segmenting the estimated F0-values into note objects using a **hidden Markow model** (**HMM**). Each note object is defined by a start time and end time in seconds, as well as an assigned F0-value (corresponding to the note's pitch). Using the interactive GUI, a user can split, merge, create or delete note objects. Finally, annotated F0-trajectories and note objects can be exported in a variety of text formats, including CSV and TXT. Tony does not offer functionalities to account for pitch drifts in performances.

The increasing scientific interest in "non-Western" music traditions [7, 9, 15, 26, 28, 43, 47, 49, 50] has led to the development of tools designed for processing and analyzing music in tuning systems other than 12-tone equal temperament (12-TET). One prominent example is Tarsos [44], a platform for analyzing pitch inventories and musical scales. Its GUI offers interactive sonifications and visualizations as well as sliders to control the included computational tools. After loading an audio file, the tool automatically computes an F0-trajectory (using the YIN algorithm [5] by default) and a pitch histogram. As opposed to Tony, Tarsos does not include functionalities to correct F0-estimates. However, it includes a tuneable "steady state filter" which allows a user to remove pitch slides in F0-trajectories. F0-trajectories and pitch histograms can be exported to different text formats and images. As for Tony, Tarsos lacks the functionality to compensate for pitch drifts in performances which is essential for our Zär scenario.

3 ZÄR DATASET

In the following, we describe our Zär dataset consisting of multitrack recordings (Section 3.1), as well as F0-annotations (Section 3.2), stable note events (Section 3.3), and pitch drift annotations (Section 3.4). We also discuss how we compute pitch inventories from the annotations (Section 3.5).

3.1 Multitrack Recordings

Our Zär dataset is based on a subset of recordings from the GVM collection [42], a collection of field recordings obtained during a research expedition in Georgia in 2016. The GVM collection comprises videos and multitrack recordings of 216 performances by different vocal ensembles. Besides a portable audio recorder that captured the whole performance, individual singers were recorded using close-up headsets and throat (larynx) microphones. Throat microphones capture the singing voice directly at the singer's throat, resulting in clean recordings with little cross-talk of other voices or other environmental sounds. In previous studies, throat microphones have shown to be beneficial for the analysis of vocal music since they nicely capture the F0 of the singer's voice [32, 37]. More information on the recording setup can be found in [42]. The 11 Zär performances (GVM-IDs 198–208) constitute a small subset of the GVM-collection. We include the multitrack recordings of these performances as mono WAV files with a sampling frequency of 22 050 Hz and 16 bit encoding. For all performances, the dataset contains at least three throat microphone signals of at least three singers (the recordings are named with suffixes `ALRX1M`, `ALRX2M`, and `ALRX3M`). The number of headset and room microphone signals varies for each performance. The 11 performances have a total duration of roughly 42 minutes.

3.2 F0-Annotations

In the context of previous studies [24, 25, 39], a Georgian ethnomusicologist semi-automatically annotated F0-trajectories of the three voices for all of the performances using the open-source tool Tony [18]. Subsequently, a domain expert double-checked the annotations. Figure 2(a) shows the F0-annotation for the top, middle, and bass voice of the performance with GVM-ID 201, which serves as our running example in the remainder of this article. The trajectories show that sung notes often start, end, or are continuously connected with pitch slides, which is a frequently used stylistic element of traditional Georgian music [33]. The F0-annotations are included as CSV files in our Zär dataset with a frequency resolution of 10 cents and a time resolution of 10 msec.

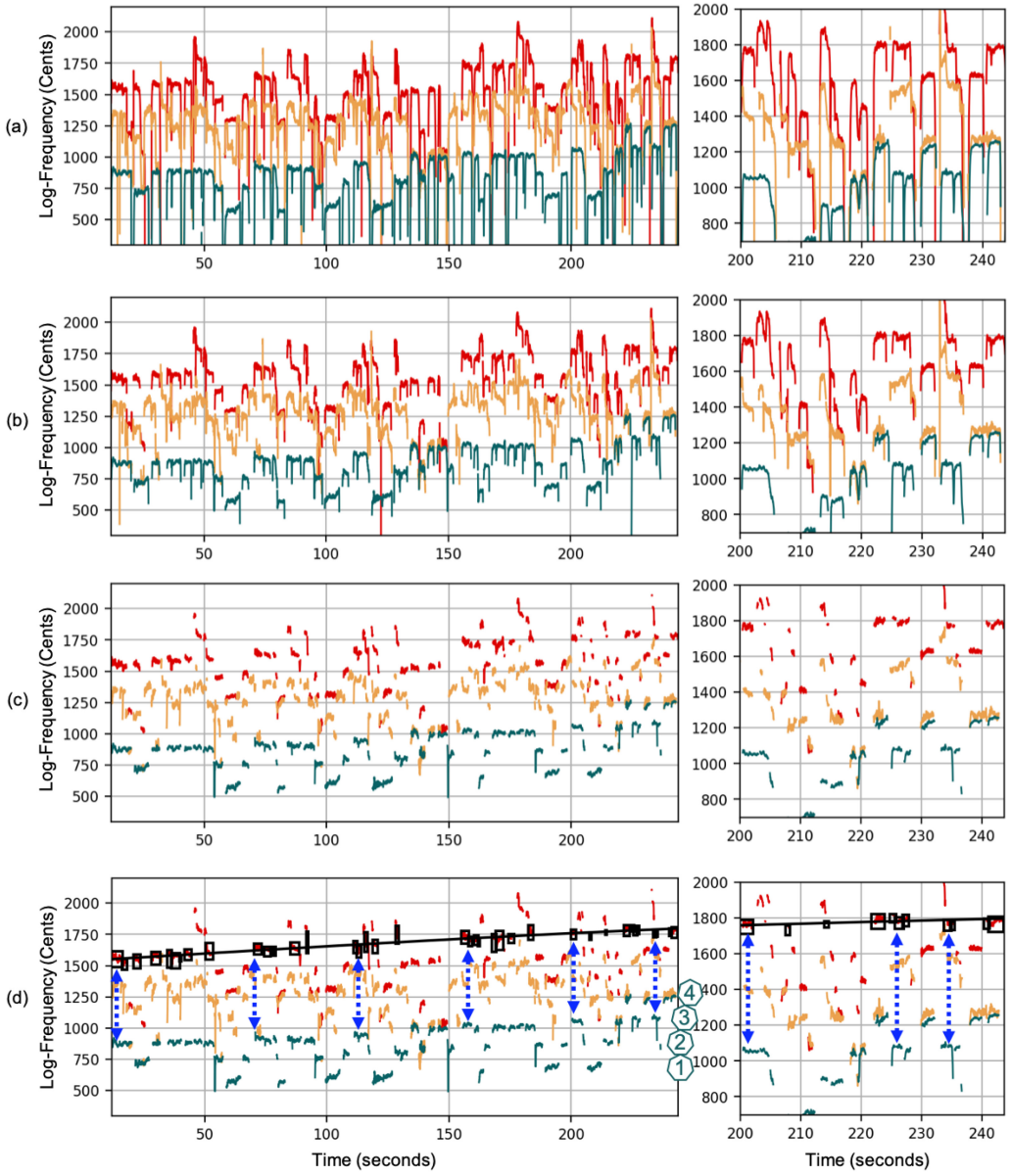

Fig. 2. Annotation process for the Zär performance with GVM-ID 201. The right column shows zoomed regions. **(a)** F0-trajectories of top, middle, and bass voice. **(b)** F0-trajectories corresponding to Tony note objects. **(c)** Stable note events. **(d)** Selected stable note events (black rectangles) and fitted drift curve (black line). The blue dashed arrows indicate parallel fifths sung by the top and bass voice.

3.3 Stable Note Events

As mentioned in Section 2, Tony automatically segments the annotated F0-trajectories into note objects. The F0-trajectories corresponding to the note objects are depicted in Figure 2(b). As one can see, Tony's automatic segmentation algorithm shortens most of the pitch slides. However, during tonal analysis or transcription of

Zär performances, the remaining slides can still lead to a significant amount of blurring and inaccuracies. One way to remove the remaining pitch slides is to use the manual correction functionalities of Tony. However, this is a time-consuming and tedious task, which is infeasible when considering larger collections. In [39], the ethnomusicologists used a heuristic to remove pitch slides by cutting off 0.15 sec at the beginning and the end of each F0-trajectory within each note object. We refer to the shortened F0-trajectories as "stable note events." The stable note events for our running example are depicted in Figure 2(c). As one can see, most of the pitch slides have been removed using this simple heuristic. In Section 4.2, we present a computational tool that helps to automize the detection of stable regions in F0-trajectories using interactive filtering techniques.

3.4 Pitch Drift Annotations

For determining the pitch drift in Zär performances, one can exploit the importance of specific harmonic intervals in traditional GVM. For instance, parallel fifths, which are often avoided in Western composed music, frequently occur in Georgian polyphonic singing [4, 35, 38]. Often, the top and bass voice sing a fifth apart (700 cents), representing the "harmonic frame" of the performance. Additionally, the unison interval (0 cents) is of great relevance in Zär performances. Essentially all songs from the region Svaneti (not only funeral dirges) end in unison. In addition, throughout a performance, the three voices of a Svan song repeatedly meet in unison, which gives the associated pitches a particular musical importance (ethnomusicologists also consider unisons as "reference pitches" in traditional Georgian singing). As a consequence, ethnomusicologists hypothesize that the pitch drift of a performance can be documented through such musically important harmonic intervals [24, 25, 39].

In [39], the ethnomusicologists followed a two-step process to determine the pitch drift in the performances. First, using a visualization as depicted in Figure 2(c), the experts visually identified a small number of stable note events in one of the voices that, according to their musical expertise, best reflect the pitch drift of the performance. For our running example, the experts selected stable note events of the top voice, which are indicated with black rectangles in Figure 2(d). As one can see, the note events were chosen to correspond to the same *scale degree* (a group of note events that roughly correspond to the same pitch after removing the drift of the performance). For instance, in Figure 2(d), the four scale degrees of the bass voice are marked using numbers in ascending order. We see that not all scale degrees are equally suitable to determine the pitch drift of the performance since some scale degrees (such as scale degree 1 of the bass voice) contain only a few stable note events. The identification of scale degrees and the selection of suitable note events require musical knowledge and need to be done with care. On closer inspection of our example, we also see that the selected note events often go along with parallel fifths of top and bass voice (see blue arrows), which shows the relevance of the fifth interval for recognizing the pitch drift of the performance.

Second, to model the pitch drift of the performance, the ethnomusicologists fitted a polynomial curve through the selected stable note events. The study in [39] revealed that polynomials of 3rd order are sufficiently suited to describe the pitch drift of the performances. Figure 2(d) shows the fitted third order polynomial (black solid line) for our example performance. Our Zär dataset includes computed drift curves with a time resolution of 10 msec as CSV files. In Section 4.3, we present a computational tool based on interactive filtering techniques for harmonic intervals and scale degrees that supports the manual selection process of note events for determining the pitch drift.

3.5 Pitch Inventories

One key towards understanding traditional Georgian tuning lies in analyzing the pitch inventories of the singers. To compute pitch inventories, the ethnomusicologists in [39] first drift-corrected the annotated stable note events with the pitch drift curve. Figure 3(a) shows the drift-corrected stable regions from Figure 2(c) using the drift curve from Figure 2(d). As one can see, the scale degrees of the three voices follow a roughly horizontal line. Subsequently, the experts computed histograms over the drift-corrected stable note events. The black line in Figure 3(b) shows the obtained max-normalized pitch inventory with a binning resolution of 10 cents. In

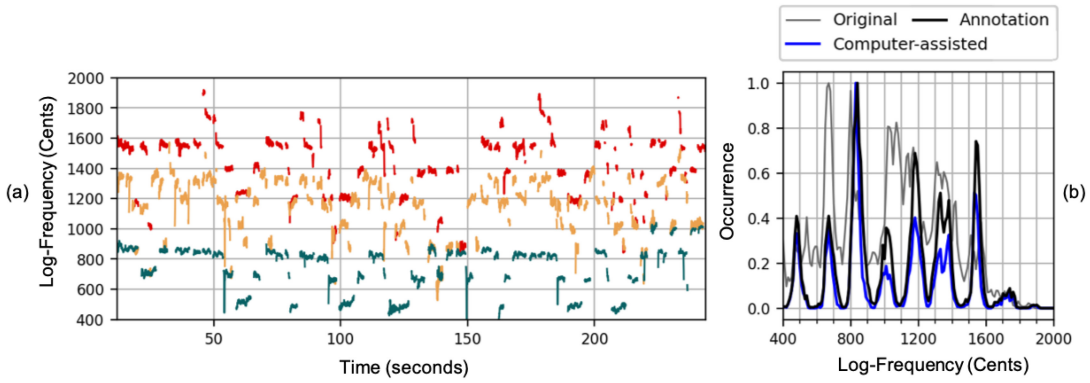

Fig. 3. Pitch inventory computation for performance with GVM-ID 201. **(a)** Drift-corrected stable note events. **(b)** Max-normalized pitch inventories.

contrast to the pitch inventory computed on the original F0-trajectories without drift correction (gray line), the annotated pitch inventory exhibits a heptatonic peak structure (seven melodic intervals per octave). The spacing between the peaks of a pitch inventory reflects the average melodic step sizes used in the performance. The pitch inventory of our running example reveals step sizes of 150–180 cents, which coincides with step sizes measured in the other Zär performances, as well as in historic recordings of liturgical chants by the former master chanter Artem Erkomaishvili [41]. Through such analysis, ethnomusicologists can obtain important cues on the tonal organization of traditional GVM. For an in-depth musicological analysis of Zär, we refer to [24, 25, 39, 40].

4 INTERACTIVE COMPUTATIONAL TOOLS

In the following, we first fix some mathematical notions (Section 4.1) and subsequently introduce our interactive computational tools for detecting stable regions (Section 4.2) and for determining pitch drift (Section 4.3).

4.1 Mathematical Notion

To account for the logarithmic nature of human pitch perception, we convert frequency values into the log-frequency domain. To this end, we fix a reference frequency ω_{ref} given in Hertz (Hz). In our experiments, we use $\omega_{\text{ref}} = 110$ Hz. Then, an arbitrary frequency value ω is converted by defining

$$F_{\text{cents}}(\omega) := 1200 \cdot \log_2\left(\frac{\omega}{\omega_{\text{ref}}}\right), \tag{1}$$

which measures the distance between ω and ω_{ref} in cents. We model an F0-trajectory as a function

$$\gamma : \mathbb{Z} \to \mathbb{R} \cup \{*\}, \tag{2}$$

which assigns to a given time index $n \in \mathbb{Z}$ either a real-valued frequency value $\gamma(n) \in \mathbb{R}$ (given in cents) or the symbol $\gamma(n) = *$ (when the frequency value is left unspecified). In our experiments, we consider trajectories with a frequency resolution of 10 cents and a time resolution of 10 msec.

4.2 Stable Region Detection

In the following, we describe a computer-assisted approach for detecting stable regions in F0-trajectories. The method is based on an existing approach using morphological (min- and max-) filters [33]. To better discuss the properties of our tool, we will recapitulate the basic steps of the approach. We explain our method using an excerpt of the top voice in the performance with GVM-ID 201. Figure 4(a) shows the given trajectory γ (black

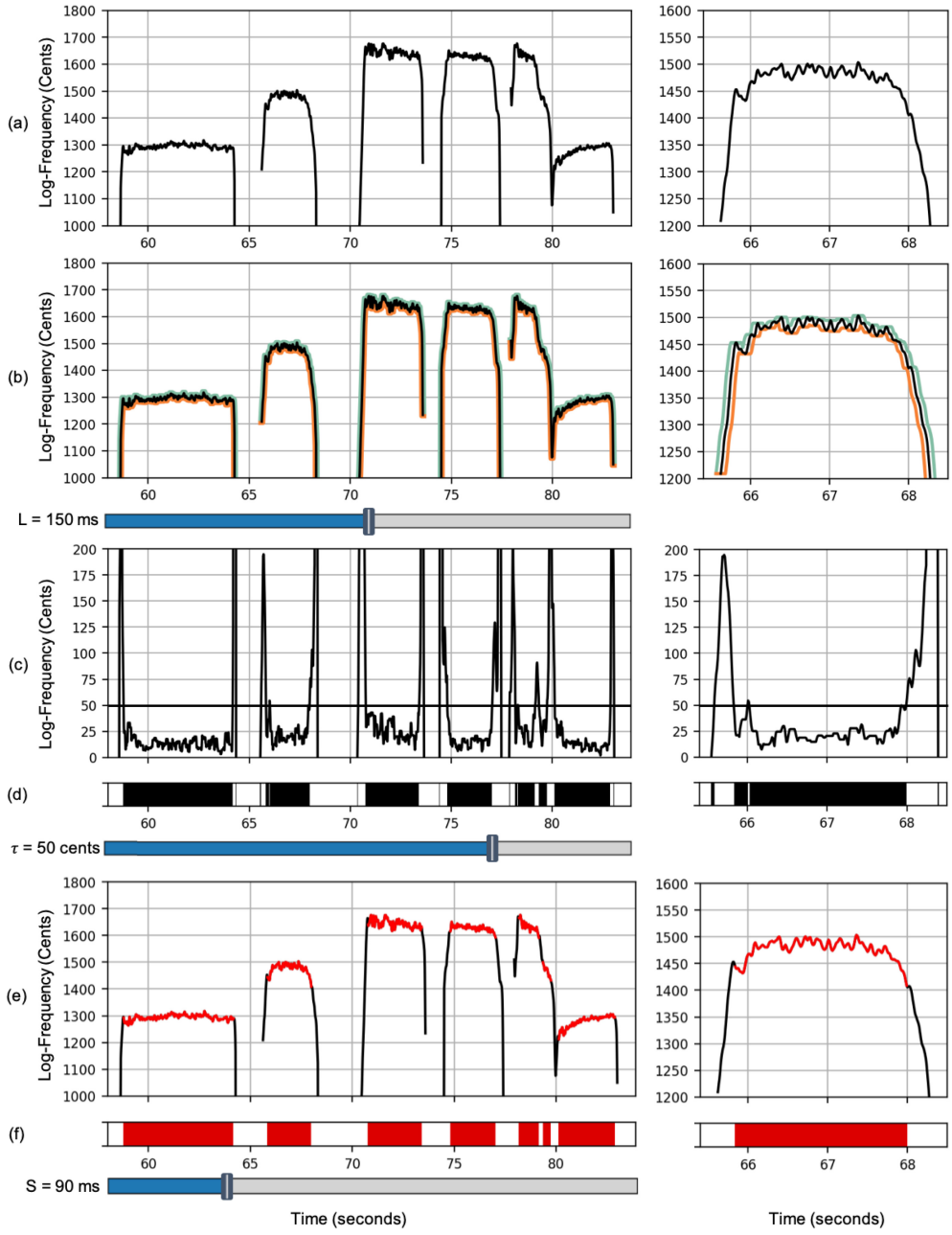

Fig. 4. Interactive detection of stable regions for the example of the top voice in performance with GVM-ID 201. The right column shows zoomed regions. **(a)** Original trajectory. **(b)** Min-filtered trajectory $\gamma^{L}_{\min}$ (orange) and max-filtered trajectory $\gamma^{L}_{\max}$ (green) for given filter length $L = 15$ (150 msec). **(c)** Morphological gradient Δ^{L} **(d)** Activation function $\mu^{L,\tau}$ after thresholding with $\tau = 50$ cents. **(e)** Trajectory γ^{Stable} restricted to stable regions (red). **(f)** Activation function $\mu^{\text{L},\tau,\text{S}}$ after smoothing with $S = 9$ (90 msec).

line), which contains several pitch slides. For brevity, we use the notion

$$\gamma(a:b) := \big\{\gamma(a), \gamma(a+1), \ldots, \gamma(b)\big\}, \tag{3}$$

for integers $a, b \in \mathbb{N}$. In a first step, we compute a dilated (max-filtered) trajectory $\gamma^L_{\max}$ and an eroded (min-filtered) trajectory $\gamma^L_{\min}$ defined by

$$\gamma^L_{\max}(n) := \max\big\{\gamma(n - \tfrac{L-1}{2} : n + \tfrac{L-1}{2})\big\}, \tag{4a}$$

$$\gamma^L_{\min}(n) := \min\big\{\gamma(n - \tfrac{L-1}{2} : n + \tfrac{L-1}{2})\big\}, \tag{4b}$$

for $n \in \mathbb{Z}$, where $L \in \mathbb{N}$ is assumed to be an odd integer. In max-filtering, the symbol $*$ is handled as $-\infty$, whereas in min-filtering it is handled as $+\infty$. Figure 4(b) shows the resulting trajectories $\gamma^L_{\min}$ (orange) and $\gamma^L_{\max}$ (green) for our running example using $L = 15$ (150 msec). In a next step, we compute the difference Δ^L between the dilated and eroded trajectories, also termed morphological gradient [31]:

$$\Delta^L(n) := \big|\gamma^L_{\max}(n) - \gamma^L_{\min}(n)\big|, \tag{5}$$

for $n \in \mathbb{Z}$, where we set $\Delta^L(n) = *$ whenever $\gamma^L_{\max}(n)$ or $\gamma^L_{\min}(n)$ are not defined. Figure 4(c) shows Δ^L for our running example. As one can see, Δ^L is large in non-stable parts (e.g., during pitch slides), whereas it is small in stable parts. After thresholding Δ^L with a chosen threshold $\tau > 0$ (given in cents), we obtain an activation function $\mu^{L,\tau}$ defined by

$$\mu^{L,\tau}(n) := \begin{cases} 1, & \text{for } \Delta^L(n) \leq \tau, \\ 0, & \text{otherwise,} \end{cases} \tag{6}$$

with $\mu^{L,\tau}(n) = 1$ indicating stable regions and $\mu^{L,\tau}(n) = 0$ indicating unstable (or undefined) regions. Figure 4(d) shows the activations $\mu^{L,\tau}$ for our running example after thresholding with $\tau = 50$ cents. As one can see, most of the stable regions of the trajectory have been correctly identified. However, there are some short passages that have been wrongly identified as stable regions (false positives). These passages result from filtering artifacts and can often be found at the beginning and end of pitch slides (e.g., at around 65.5 sec) or in between two fastly succeeding notes (e.g., at around 79 sec). Also, there are some short interruptions in stable regions (false negatives), e.g., at 66 sec.

In practice, one may want to remove these outliers and obtain coherent entities of F0-values, similar to the stable note events from Section 3.3. Therefore, as an extension to the original approach in [33], we propose an optional smoothing step of the trajectory activations $\mu^{L,\tau}$ by applying a median filter:

$$\mu^{\mathrm{L},\tau,\mathrm{S}}(n) := \mathrm{median}\big\{\mu^{L,\tau}(n - \tfrac{S-1}{2} : n + \tfrac{S-1}{2})\big\}, \tag{7}$$

where $S \in \mathbb{N}$ is assumed to be an odd integer and the symbol $*$ is handled as $-\infty$. Note that by setting $S = 1$, no smoothing is applied. In a final step, we compute the trajectory restricted to stable regions γ^{Stable} by

$$\gamma^{\mathrm{Stable}}(n) := \begin{cases} \gamma(n), & \text{for } \mu^{\mathrm{L},\tau,\mathrm{S}}(n) = 1, \\ *, & \text{otherwise.} \end{cases} \tag{8}$$

Figure 4(e) shows the resulting trajectory and its activation function after smoothing with a median filter of length $S = 9$ (90 msec). As one can see, most of the outliers have been removed (only the outlier at around 79 sec remains). One may further tackle such outliers by applying additional heuristics such as removing detected stable regions that fall below a certain minimal length or that exceed a certain variance. Note that the algorithm leaves the frequency values of the original F0-trajectory unaltered (e.g., no quantization or smoothing of frequency values), which is important for subsequent tonal analysis steps.

In practice, an ethnomusicologist (without explicit knowledge of signal processing) can use interactive visualizations similar to Figure 4 for tuning the three parameters L, τ, and S of our algorithm. The min-/max- filter

length L controls the sensitivity of the method to (sudden) fluctuations in the F0-trajectory. Small L may lead to an increased number of false positives, while large L lead to an increased number of false negatives (in particular at the beginning and end of stable regions) [33]. The threshold τ can be seen as a tolerance parameter that specifies the maximally allowed fluctuation under which a trajectory is still considered to be stable. Therefore, τ may be tuned according to the singing style (e.g., the amount of vibrato) used in the performance or the singing proficiency. After determining L and τ, the smoothing filter of length S can be tuned to refine the detection by removing outliers observed in the activation function $\mu^{L,\tau}$, which results in stable, note-like events. In summary, the three parameters of our tool have an explicit and easy-to-understand meaning, which is important for use in interdisciplinary research.

4.3 Drift Estimation

In the following, we describe a computer-assisted approach for estimating the pitch drift of a Zär performance using interactive filtering techniques for harmonic intervals and scale degrees. Our approach is based on the hypothesis that certain (musically important) harmonic intervals capture the pitch drift of a performance (see Section 3.4). We explain our method along with Figure 5, using again the performance with GVM-ID 201 as an example. In the following, we assume a set of M trajectories

$$\Gamma := \{\gamma_1, \ldots, \gamma_M\}, \tag{9}$$

where γ_m is the F0-trajectory of the mth voice, $m \in [1 : M]$. In our example, we consider F0-trajectories restricted to stable regions for the top, middle, and bass voice ($M = 3$) as depicted in Figure 5(a). As one can see, the three singers continuously drift upwards over the course of the performance.

As discussed in Section 3.4, the pitch drift of a Zär performance is captured by certain musically important harmonic intervals (e.g., the unison or the fifth). Therefore, in the first step, we filter the given F0-trajectories with respect to a user-specified harmonic interval. For a given $m \in [1 : M]$ and an interval I in cents, let $\mathcal{H}_m^I$ denote the set that contains all time indices n for which there is at least one other trajectory γ_k, $k \in [1 : M] \setminus \{m\}$ that is I cents apart up to a tolerance ε in cents. In other words:

$$\mathcal{H}_m^I := \Big\{ n \in \mathbb{Z} | \exists k \in [1 : M] \setminus \{m\} : I - \varepsilon \leq |\gamma_m(n) - \gamma_k(n)| \leq I + \varepsilon \Big\}. \tag{10}$$

We then define the interval-filtered F0-trajectory γ_m^I for voice m by

$$\gamma_m^I(n) := \begin{cases} \gamma_m(n), & \text{for } n \in \mathcal{H}_m^I, \\ *, & \text{otherwise.} \end{cases} \tag{11}$$

Figure 5(b) illustrates our interval-filtering for the fifth interval ($I = 700$ cents) with $\varepsilon = 20$ cents for the top and bass voice ($M = 2$). It can be seen that the remaining F0-values of the top and bass voice are spaced roughly 700 cents apart (as indicated by the blue dotted arrows).

Similar to Section 3.4, in the next step, the user selects a scale degree that best reflects the pitch drift of the performance. In our example, the user chooses the third scale degree of the bass voice (counting the remaining scale degrees after interval filtering from low to high). The F0-values corresponding to the chosen scale can be obtained using a suitable clustering algorithm.

In our work, we use a simple two-step clustering method inspired by the Radon Transform [29]: first, we rotate the interval-filtered trajectories around the coordinate origin such that the entropy of a computed pitch inventory is minimized. The entropy indicates the peakedness of a distribution while being low for peaked distributions and high for flat distributions. Thus, the rotation angle that minimizes entropy constitutes an approximation of the linear drift slope of the performance. This entropy-minimizing rotation angle can be determined automatically through an exhaustive search over a musically meaningful range of angles. In our experiments, we assume that the singers do not drift more than $\pm 1,200$ cents (or an octave) over the course of a performance, which is a reasonable choice for Zär performances [24, 39]. Second, we perform k-means clustering on the interval-filtered

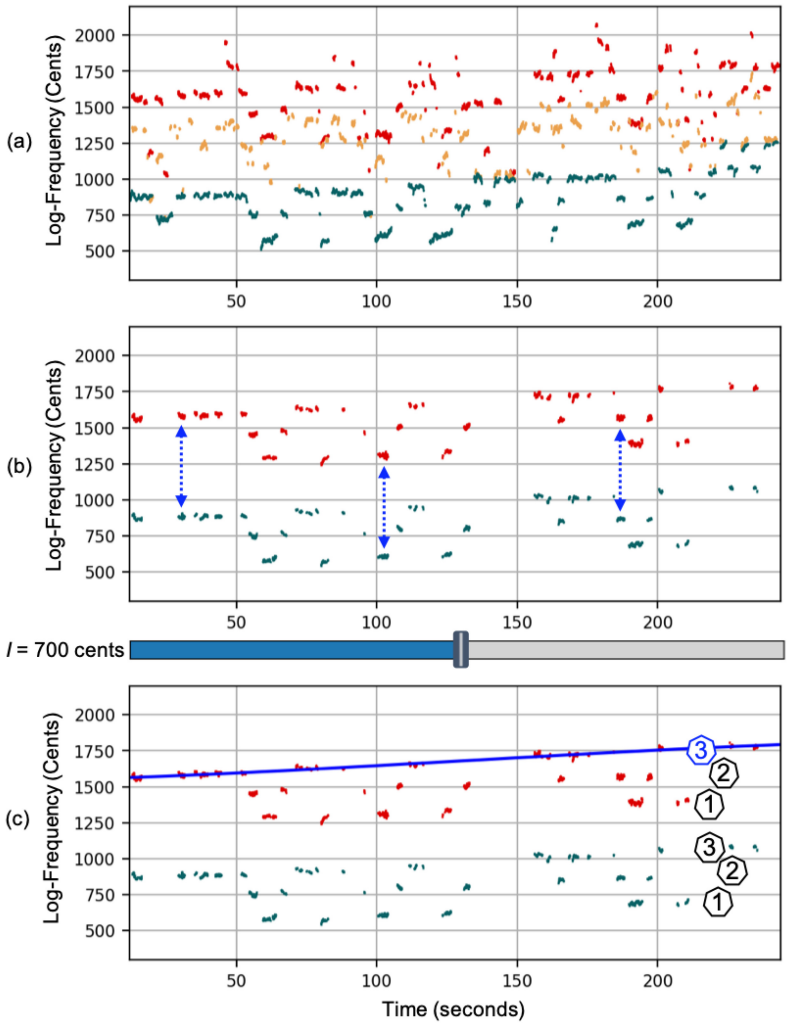

Fig. 5. Interactive drift estimation illustrated by the performance with GVM-ID 201. **(a)** Stable note events. **(b)** Interval-filtered F0-trajectories with interval $I = 700$ cents using the top and bass voice ($M = 2$). **(c)** Polynomial fitting using the F0-values of a manually specified scale degree.

and rotated trajectories, with k corresponding to the number of scale degrees of each voice ($k = 3$ in our example). As in Section 3.4, we fit a polynomial of third order using the F0-values that correspond to the chosen scale degree. The resulting drift curve for our example is indicated by the blue line in Figure 5(c). Note that polynomial fitting through a certain scale degree of the bass voice should result in a similar drift curve than polynomial fitting through the same scale degree of the top voice.

In practice, using our tool and visualizations similar to Figure 5, an ethnomusicologist can interactively explore and analyze how different harmonic intervals I and scale degrees reflect the pitch drift of the performance. An additional indicator for the correctness of a determined pitch drift are pitch inventories. The blue line in Figure 3(b) shows the pitch inventory of our running example obtained through drift-correcting the detected stable regions from Figure 5(a) with the normalized drift curve from Figure 5(c). As one can see, the computed pitch inventory has a similar peak structure to the annotated pitch inventory, which indicates that the pitch

drift has correctly been determined. The peaks at around 1,000, 1,200, and 1,400 cents are less prominent in the computed pitch inventory compared to the annotated pitch inventory. This observation indicates that for our example and chosen parameter settings, there remain fewer F0-values after stable region detection than in the stable region annotation (compare Figure 2(c) with Figure 3(a)).

5 COMPUTER-ASSISTED ANALYSIS OF ZÄR PERFORMANCES

In this section, we discuss how our interactive computational tools can be applied to support ethnomusicological research. First, in a case study of Georgian Zär, we use our tools to reproduce pitch inventories of a previous study (Section 5.1). Second, in the light of our experimental results, we discuss the potential of computational tools for ethnomusicological research (Section 5.2).

5.1 A Case Study of Pitch Inventories

We now show how a domain expert can use our computational tools to reproduce the pitch inventories from the previous study on Georgian Zär [39]. Since the musical meaning of pitch inventories is hard to quantify and evaluate, we discuss different qualitative aspects of our work along with the visualizations in Figure 6. Figure 6(a) shows the annotated stable note events and pitch drifts (black lines) from our Zär dataset (Section 3.4) as reference.

In our case study, we start with the F0-annotations of the three voices described in Section 3.2. Note that in case no F0-annotations are at hand, one can use automatic approaches such as the one proposed in [34] to obtain reliable F0-estimates. In a first step, we use the tool introduced in Section 4.2 to determine stable regions in the F0-trajectories. Using interactive visualizations such as Figure 4, a domain expert can tune parameters L, τ, and S such that the pitch slides are removed. In our study, we set $L = 15$ bins (150 msec), which corresponds to the value that the domain experts chose for determining stable note events from Section 3.3. Furthermore, we set $\tau = 50$ cents, which is a reasonable value for Georgian singing. To refine the detection, we empirically determined $S = 9$ bins (90 msec), which removes short outliers. Finally, we remove stable regions that are shorter than 100 msec to further refine the detection. The resulting trajectories restricted to stable regions are depicted in Figure 6(b). Overall, the filtered trajectories resemble the manually annotated stable note events from Figure 6(a).

In a second step, we use the tool introduced in Section 4.3 to determine the pitch drifts of the performances. Using musical domain knowledge and interactive visualizations such as Figure 5, an expert can choose an interval I and a scale degree that best capture the pitch drift of a performance. As explained in Section 3.5, the unison and the fifth interval are of special musical importance in Georgian Zär. For the performances with GVM-ID 199–206, which exhibit very prominent parallel fifths, we filtered for the fifth interval ($I = 700 \pm 20$ cents) of top and bass voice. For the performances with GVM-ID 198, 207, and 208, which exhibit less prominent parallel fifths, we chose to filter for the unison interval ($I = 0 \pm 20$ cents) considering all voices. The interval filtered trajectories of all performances are depicted in Figure 6(c). Subsequently, we use the clustering algorithm described in Section 4.3 to automatically determine the scale degrees of the interval filtered trajectories. For the performances 199–206, we determined $k = 6$ clusters (3 for each voice), and for the performances 198, 207, and 208, we determined $k = 3$ clusters. In our case study, we selected similar scale degrees as the domain expert in the manual study. The fitted polynomial drift curves through the selected scale degrees are shown as blue lines in Figure 6(c). As one can see, the drift curves have a similar progression compared to the annotated pitch drifts in Figure 6(a). Note that instead of advocating a specific interval or scale degree, this case study shows only one way how the pitch drift in Zär performances can be determined. For instance, our computational tools enable domain experts to explore interval filtering for different harmonic intervals as well as suitable combinations, which may lead to more accurate drift estimates. We leave an investigation of these aspects for future work.

In a final step, we compensate for the pitch drift of the trajectories restricted to stable regions from Figure 6(b) with the normalized drift curves from Figure 6(c) before computing pitch inventories. In our experiments, we

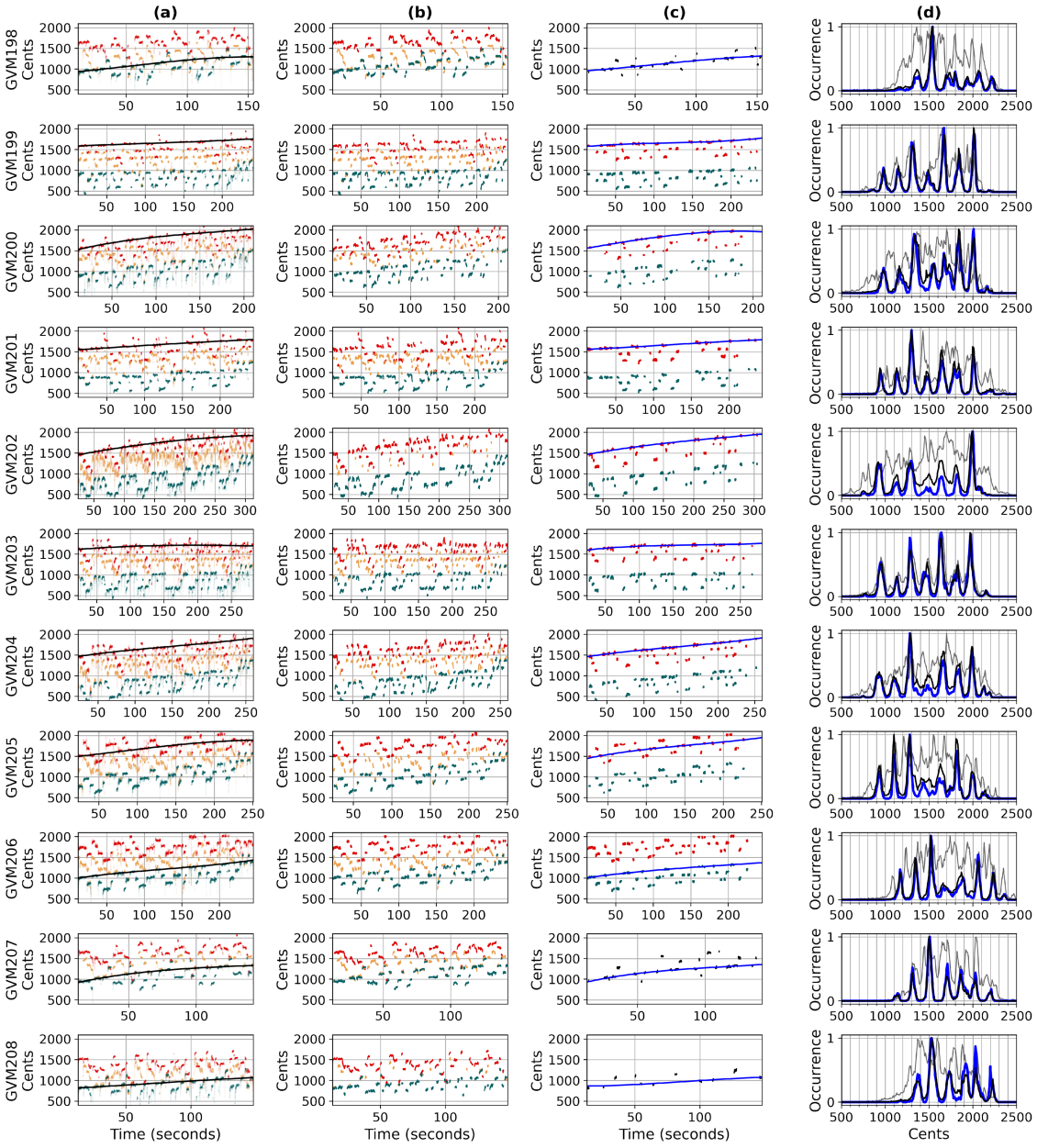

Fig. 6. Pitch inventory computation for all 11 Zär performances. **(a)** Manually annotated stable note events and reference pitch drift. **(b)** Trajectories are restricted to stable regions using the interactive tool from Section 4.2. **(c)** Interval-filtered trajectories and estimated pitch drift using the interactive tool from Section 4.3. **(d)** Pitch inventories (for legend, see Figure 3(b)).

use a binning resolution of 10 cents and max-normalize all pitch inventories. Note that for tonal analysis, one is mainly interested in the relative peak positions of pitch inventories (see Section 3.5). In order to facilitate the visual comparison of pitch inventories across performances, we shift all F0-values by a constant amount (which differs for each Zär) in such a way that the final long note (with a duration of at least 1 sec) in the middle voice has a pitch of 1,500 cents. Figure 6(d) shows the pitch inventories obtained from the original F0-trajectories without drift correction (gray), the pitch inventories based on the reference annotations (black), and the pitch inventories obtained using our computer-assisted tools (blue). We can see that the pitch inventories computed with the help of our interactive tools are very similar to the reference pitch inventories.

5.2 Applications to Computational Ethnomusicology

The study described in Section 3 exemplifies many of the challenges that ethnomusicological studies on field recordings face. Even relatively basic analysis tasks, such as the computation of pitch inventories, typically require multiple annotation steps conducted by domain experts. State-of-the-art annotation tools such as Tony or Praat face several limitations since they are either designed for "Western" music or lack necessary functionalities. This often leads to labor-intensive annotation processes with tedious manual correction steps. While these efforts were made for the 11 performances of our Zär dataset, similar studies on larger corpora, such as the whole GVM collection [42], would be very time-consuming to perform. Also, such highly manual annotation and analysis processes can suffer from subjective decisions, thus making it hard to reproduce the results.

As our case study showed, computational tools can support ethnomusicological studies on field recordings by taking over specific, well-defined tasks of the annotation process under the guidance of a domain expert. Through tuning a few musically motivated parameters and suitable interactive visualizations, a domain expert could reproduce the pitch inventories that were obtained by tedious manual annotations in significantly less time. In this way, our computer-assisted procedure can accelerate and simplify musicological analyses as well as enable the exploration of large music corpora to gain new musicological insights.

6 CONCLUSIONS

In this article, we presented a publicly available dataset based on 11 performances of three-voice Georgian Zär, which includes expert annotations of F0-trajectories, stable note events, and pitch drifts. The dataset is of high value for ethnomusicological research and the preservation of the Georgian musical heritage. Furthermore, we introduced two computational tools based on interactive filtering techniques for detecting stable regions in F0-trajectories and determining the pitch drift of the performances. In a case study of pitch inventories of Zär performances, we showed that our computational tools can help to make ethnomusicological research on Georgian Zär and possibly other non-Western singing traditions more efficient. Furthermore, our tools open up new ways to explore data collections. In future work, we plan to use our computer-assisted approaches in an exploratory study on the entire GVM collection.

ACKNOWLEDGMENTS

We thank Nana Mzhavanadze for contributing to the annotations of the Zär dataset.

REFERENCES

[1] Per-Gunnar Alldahl. 2008. *Choral Intonation*. Gehrmans Musikförlag.
[2] Rytis Ambrazevičius, Robertas Budrys, and Irena Višnevska. 2015. *Scales in Lithuanian Traditional Music: Acoustics, Cognition, and Contexts*. Arx Reklama.
[3] Chris Cannam, Christian Landone, and Mark B. Sandler. 2010. Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the International Conference on Multimedia*. Florence, Italy, 1467–1468.
[4] Evsevi Chokhonelidze. 2010. Some characteristic features of the voice coordination and harmony in Georgian multipart singing. In *Proceedings of the Echoes from Georgia: Seventeen Arguments on Georgian Polyphony*. Nova Science Publishers, 135–145.

[5] Alain de Cheveigné and Hideki Kawahara. 2002. YIN, a fundamental frequency estimator for speech and music. *Journal of the Acoustical Society of America* 111, 4 (2002), 1917–1930.

[6] Malkhaz Erkvanidze. 2016. The Georgian musical system. In *Proceedings of the International Workshop on Folk Music Analysis*. Dublin, Ireland, 74–79.

[7] Kaustuv Kanti Ganguli and Preeti Rao. 2018. On the distributional representation of ragas: Experiments with allied raga pairs. *Transactions of the International Society for Music Information Retrieval* 1, 1 (2018), 79–95. DOI: https://doi.org/10.5334/tismir.11

[8] Ali C. Gedik and Barış Bozkurt. 2010. Pitch-frequency histogram-based music information retrieval for Turkish music. *Signal Processing* 90, 4 (2010), 1049–1063.

[9] Emilia Gómez, Perfecto Herrera, and Francisco Gómez-Martin. 2013. Computational ethnomusicology: Perspectives and challenges. *Journal of New Music Research* 42, 2 (2013), 111–112. DOI: https://doi.org/10.1080/09298215.2013.818038

[10] David M. Howard. 2007. Intonation drift in a capella soprano, alto, tenor, bass quartet singing with key modulation. *Journal of Voice* 21, 3 (2007), 300–315. DOI: https://doi.org/10.1016/j.jvoice.2005.12.005

[11] Nino Kalandadze–Makharadze. 2004. The funeral Zari in traditional male polyphony. In *Proceedings of the International Symposium on Traditional Polyphony*. Tbilisi, Georgia, 166–178.

[12] Richard Keeling. 1985. Contrast of song performance style as a function of sex role polarity in the Hupa Brush Dance. *Ethnomusicology* 29, 2 (1985), 185–212.

[13] Gopala Krishna Koduri, Sankalp Gulati, Preeti Rao, and Xavier Serra. 2012. Rāga recognition based on pitch distribution methods. *Journal of New Music Research* 41, 4 (2012), 337–350. DOI: https://doi.org/10.1080/09298215.2012.735246

[14] Gopala Krishna Koduri, Joan Serrà, and Xavier Serra. 2012. Characterization of intonation in carnatic music by parametrizing pitch histograms. In *Proceedings of the International Society for Music Information Retrieval Conference*. Porto, Portugal, 199–204. DOI: https://doi.org/10.5281/zenodo.1416902

[15] Nadine Kroher and Emilia Gómez. 2016. Automatic transcription of flamenco singing from polyphonic music recordings. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 24, 5 (2016), 901–913. DOI: https://doi.org/10.1109/TASLP.2016.2531284

[16] Jiei Kuroyanagi, Shoichiro Sato, Meng-Jou Ho, Gakuto Chiba, Joren Six, Peter Pfordresher, Adam Tierney, Shinya Fujii, and Patrick Savage. 2019. Automatic comparison of human music, speech, and bird song suggests uniqueness of human scales. In *Proceedings of the Folk Music Analysis Conference*. 35–40.

[17] Edith L. M. Law, Luis von Ahn, Roger B. Dannenberg, and Mike Crawford. 2007. TagATune: A game for music and sound annotation. In *Proceedings of the International Society for Music Information Retrieval Conference*. Vienna, Austria, 361–364. DOI: https://doi.org/10.5281/zenodo.1415568

[18] Matthias Mauch, Chris Cannam, Rachel Bittner, George Fazekas, Justing 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*.

[19] Matthias Mauch and Simon Dixon. 2014. pYIN: A fundamental frequency estimator using probabilistic threshold distributions. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Florence, Italy, 659–663. DOI: https://doi.org/10.1109/ICASSP.2014.6853678

[20] Mervyn McLean. 1964. A preliminary analysis of 87 Maori chants. *Ethnomusicology* 8, 1 (1964), 41–48.

[21] Blai Meléndez-Catalán, Emilio Molina, and Emilia Gómez. 2017. BAT: An open-source, web-based audio events annotation tool. In *Proceedings of the Web Audio Conference*.

[22] Dirk Moelants, Olmo Cornelis, and Marc Leman. 2009. Exploring African tone scales. In *Proceedings of the International Society for Music Information Retrieval Conference*. Kobe, Japan, 489–494. DOI: https://doi.org/10.5281/zenodo.1416338

[23] Meinard Müller, 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*. Erlangen, Germany, 186–193.

[24] Nana Mzhavanadze and Frank Scherbaum. 2020. Svan funeral dirges (Zär): Musicological analysis. *Musicologist* 4, 2 (2020), 168–197. DOI: https://doi.org/10.33906/musicologist.782185

[25] Nana Mzhavanadze and Frank Scherbaum. 2021. Svan funeral dirges (Zär): Cultural context. *Musicologist* 5, 2 (2021), 133–165. DOI: https://doi.org/10.33906/musicologist.906765

[26] Yuto Ozaki, John M. McBride, Emmanouil Benetos, Peter Pfordresher, Joren Six, Adam Tierney, Polina Proutskova, Emi Sakai, Haruka Kondo, Haruno Fukatsu, Shinya Fujii, and Patrick E. Savage. 2021. Agreement among human and automated transcriptions of global songs. In *Proceedings of the International Society for Music Information Retrieval Conference*. 500–508. DOI: https://doi.org/10.5281/zenodo.5624529

[27] Sachin Pant, Vishweshwara Rao, and Preeti Rao. 2010. A melody detection user interface for polyphonic music. In *Proceedings of the National Conference on Communications*. Chennai, India, 1–5.

[28] Maria Panteli, Emmanouil Benetos, and Simon Dixon. 2018. A review of manual and computational approaches for the study of world music corpora. *Journal of New Music Research* 47, 2 (2018), 176–189. DOI: https://doi.org/10.1080/09298215.2017.1418896

[29] John Radon. 1917. Über die Bestimmung von Funktionen längs gewisser Mannigfaltigkeiten. Berichte über die Verhandlungen der Königlich-Sächsischen Gesellschaft der Wissenschaften zu Leipzig. *Mathematisch-Physische Klasse* 69 (1917), 262–277.
[30] António Ramires, Frederic Font, Dmitry Bogdanov, Jordan B. L. Smith, Yi-Hsuan Yang, Joann Ching, Bo-Yu Chen, Yueh-Kao Wu, Hsu Wei-Han, and Xavier Serra. 2020. The freesound loop dataset and annotation tool. In *Proceedings of the International Society for Music Information Retrieval Conference*. Montreal, Canada, 287–294. DOI: https://doi.org/10.5281/zenodo.4245430
[31] Jean-François Rivest, Pierre Soille, and Serge Beucher. 1993. Morphological gradients. *Journal of Electronic Imaging* 2, 4 (1993), 326–336.
[32] Sebastian Rosenzweig, Helena Cuesta, Christof Weiß, Frank Scherbaum, Emilia Gómez, and Meinard Müller. 2020. Dagstuhl ChoirSet: A multitrack dataset for MIR research on choral singing. *Transactions of the International Society for Music Information Retrieval* 3, 1 (2020), 98–110. DOI: https://doi.org/10.5334/tismir.48
[33] Sebastian Rosenzweig, 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*. Delft, The Netherlands, 352–359. DOI: https://doi.org/10.5281/zenodo.3527816
[34] Sebastian Rosenzweig, Frank Scherbaum, and Meinard Müller. 2021. Reliability assessment of singing voice F0-estimates using multiple algorithms. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*. Toronto, Canada, 261–265. DOI: https://doi.org/10.1109/ICASSP39728.2021.9413372
[35] Sebastian Rosenzweig, 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* 3, 1 (2020), 31–41. DOI: https://doi.org/10.5334/tismir.44
[36] Shoichiro Sato, Joren Six, Peter Pfordresher, Shina Fujii, and Patrick Savage. 2019. Automatic comparison of global children's and adult songs supports a sensorimotor hypothesis for the origin of musical scales. In *Proceedings of the International Workshop on Folk Music Analysis*. Birmingham, UK, 41–46.
[37] Frank Scherbaum. 2016. On the benefit of larynx-microphone field recordings for the documentation and analysis of polyphonic vocal music. *Proceedings of the International Workshop Folk Music Analysis* (2016), 80–87.
[38] Frank Scherbaum, Meinard Müller, and Sebastian Rosenzweig. 2017. Analysis of the Tbilisi State Conservatory recordings of Artem Erkomaishvili in 1966. In *Proceedings of the International Workshop on Folk Music Analysis*. Málaga, Spain, 29–36.
[39] Frank Scherbaum and Nana Mzhavanadze. 2020. Svan funeral dirges (Zär): Musical acoustical analysis of a new collection of field recordings. *Musicologist* 4, 2 (2020), 138–167. DOI: https://doi.org/10.33906/musicologist.782094
[40] Frank Scherbaum and Nana Mzhavanadze. 2021. Svan funeral dirges (Zär): Language-music relation and phonetic properties. *Musicologist* 5, 1 (2021), 66–82. DOI: https://doi.org/10.33906/musicologist.875348
[41] Frank Scherbaum, Nana Mzhavanadze, Simha Arom, Sebastian Rosenzweig, and Meinard Müller. 2020. *Tonal Organization of the Erkomaishvili Dataset: Pitches, Scales, Melodies and Harmonies*. Universitätsverlag Potsdam. DOI: https://doi.org/10.25932/publishup-47614
[42] Frank Scherbaum, Nana Mzhavanadze, Sebastian Rosenzweig, and Meinard Müller. 2019. Multi-media recordings of traditional Georgian vocal music for computational analysis. In *Proceedings of the International Workshop on Folk Music Analysis*. Birmingham, UK, 1–6.
[43] Xavier Serra. 2014. Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the AES International Conference on Semantic Audio*. London, UK.
[44] Joren Six, Olmo Cornelis, and Marc Leman. 2013. Tarsos, a modular platform for precise pitch analysis of Western and Non-Western music. *Journal of New Music Research* 42, 2 (2013), 113–129. DOI: https://doi.org/10.1080/09298215.2013.797999
[45] Johan Sundberg. 1987. *The Science of the Singing Voice*. Northern Illinois University Press.
[46] Zaal Tsereteli and Levan Veshapidze. 2014. On the Georgian traditional scale. *Proceedings of the International Symposium on Traditional Polyphony*. 288–295.
[47] George Tzanetakis. 2014. Computational ethnomusicology: A music information retrieval perspective. In *Proceedings of the Joint Conference 40th International Computer Music Conference and 11th Sound and Music Computing Conference*. Athens, Greece, 69–73.
[48] George Tzanetakis, Andrey Ermolinskyi, and Perry Cook. 2003. Pitch histograms in audio and symbolic music information retrieval. *Journal of New Music Research* 32, 2 (2003), 143–152. DOI: https://doi.org/10.1076/jnmr.32.2.143.16743
[49] George Tzanetakis, Ajay Kapur, W. Andrew Schloss, and Matthew Wright. 2007. Computational ethnomusicology. *Journal of Interdisciplinary Music Studies* 1, 2 (2007), 1–24.
[50] Peter van Kranenburg, Martine de Bruin, and Anja Volk. 2019. Documenting a song culture: The Dutch Song Database as a resource for musicological research. *International Journal on Digital Libraries* 20, 1 (2019), 13–23.

Received 12 November 2021; revised 6 May 2022; accepted 6 May 2022