TEACH YOURSELF GEORGIAN FOLK SONGS DATASET: AN ANNOTATED CORPUS OF TRADITIONAL VOCAL POLYPHONY

David Gillman
New College of Florida
dgillman@ncf.edu

Uday Goyat
Georgia Institute of Technology
ugoyat3@gatech.edu

Atalay Kutlay
New College of Florida
atalay.kutlay18@ncf.edu

ABSTRACT

New datasets of non-Western traditional music contribute to the development of knowledge in MIR and allow computational techniques to inform ethnomusicology. We present an annotated dataset of traditional vocal polyphony from two regions of the Republic of Georgia with disparate musical characteristics. The audio for each song consists of four polyphonic recordings of one performance from different microphones. We present a process and workflow that we use to annotate the dataset, which takes advantage of the salience of individual voices in each recording. The process results in an f_0 estimate for each vocal part.

1. INTRODUCTION

To evaluate algorithms in Music Information Retrieval (MIR) it is essential to have a variety of extensive datasets of annotated music [1]. Annotated datasets of vocal *a cappella* music have been scarce until recently, particularly of non-Western music. The work of [2] provides a list of datasets of vocal polyphony that includes two datasets of songs from the Republic of Georgia: the Erkomaishvili dataset of [3] and the collection recorded in the work of [4].

Multi-f_0 estimation is a sub-problem of Automated Music Transcription (AMT) consisting of identifying the fundamental frequency f_0 of each part in a polyphonic recording. Datasets of polyphony that are annotated with the fundamental frequency of each part are useful as ground truth for multi-f_0 algorithms. Multi-f_0 estimation is a challenging problem for *a cappella* vocal music because of the variety of sounds produced by the human voice and the similarity in timbre of different voices [5].

In this work we present an annotated dataset of 38 three-part songs from the Republic of Georgia, including 29 from the region of Guria and nine from the region of Samegrelo. The total duration of the collection is 89 minutes. Gurian songs are a particular challenge for multi-f_0 estimation. The three parts are independent melodic lines that often cross and contain rapid movement. The top part often consists of *krimanchuli*, Georgian yodeling, with as many as six changes of octave per second. These features make our dataset a useful contribution as part of a training set for multi-f_0 algorithms. We have created a web-based visualization of the dataset that is potentially useful as an aid to singers learning their parts on Georgian songs, which was the purpose of the original recordings.

We also present a new process and workflow for multi-f_0 estimation where several recordings exist of a single performance. Our dataset is an unusual challenge for annotation in that it does not include isolated tracks for each vocal part. Instead there are four recordings of one performance made from different microphones. One recording presents a balanced mix of voices. In each of the other three recordings one of the voices is more salient than the other two, but all three voices are easily audible and create a polyphonic mixture. Our process and workflow consist of isolating the salient voice, applying several algorithms for monophonic f_0 estimation, and using a graphical interface to select the correct estimate. In addition to f_0 estimates we present the median absolute deviation of the estimates, a measure of confidence in the estimates.

Other interactive methods have appeared for extracting melody from audio. The work of [6] introduced the Tony software for monophonic audio. It presents to the user several pitch estimates generated by an early stage of the pYin algorithm, from which the user can select ranges of time and frequency. It is designed for ease of use and has many features such as octave correction [7]. The system of [8] designed for the Erkomaishvili dataset presents to the user a spectrogram with one melody highlighted as a result of dynamic programming performed on a set of salient frequencies at each time step, followed by automatic corrections that take into account musical knowledge such as voice ranges. The user is able to delete and replace lines in the spectrogram. There is a web interface for the Erkomaishvili dataset which plays the audio of each song accompanied by a scrolling score with lyrics. The work of [2] created the Dagstuhl dataset by recording each singer with a larynx microphone, a headset microphone, and a dynamic microphone. The researchers applied both the pYin and CREPE algorithms to each recording and derived confidences for each algorithm on each microphone using a subset of recordings manually annotated by a sound engineer who used Tony.

Our method differs from these methods in that it is designed for polyphonic recordings that contain one salient voice, it makes use of several f_0 estimates, and it presents

Attribution: D. Gillman, U. Goyat, and A. Kutlay, "*Teach Yourself Georgian Folk Songs* Dataset: An Annotated Corpus Of Traditional Vocal Polyphony", in *Proc. of the 23rd Int. Society for Music Information Retrieval Conf.,* Bengaluru, India, 2022.

a measure of confidence in the estimates. We have created a web interface for the dataset which plays the audio of each song accompanied by a scrolling visualization of the pitches. Unlike that of [2] our visualization shows the f_0 estimates of the three vocal parts, and the median absolute deviation of the each estimate is displayed in the manner of a confidence interval.

1.1 Georgian music datasets

The Republic of Georgia, bounded on the north by the Caucasus Mountains and on the west by the Black Sea, contains within its borders several starkly varying traditions of three-part *a cappella* vocal music [9]. Georgian singing is an aural tradition that uses a scale of its own [10]. The intervals that make up the Georgian scale are an area of research in ethnomusicology, as well as the extent to which singers in the tradition adjust to one another and deviate from a fixed scale to produce desired harmonies [11, 12]. Annotated datasets of Georgian music have been useful in advancing this research [9, 13].

The Erkomaishvili dataset contains Georgian sacred songs performed by a single performer on all three parts, with overdubbing. The most recent version of the dataset due to [3] includes f_0 annotations due to [8], onset annotations, and musical notation due to [14]. The dataset due to [4] includes Georgian folk songs from various regions recorded using individual larynx and headset microphones and a microphone for the ensemble. The authors observe that larynx microphones isolate the three parts well, and they illustrate this point by showing estimates of f_0 derived for one song in the collection.

The present dataset, like these other datasets, consists of recordings of performances by expert Georgian singers of interest to ethnomusicologists. However, unlike the dataset due to [3], our dataset consists of studio recordings of live ensembles, and unlike the dataset due to [4], our dataset was recorded without the benefit larynx microphones and therefore presents a greater challenge for pitch estimation.

Our dataset contains the recordings in the collections *Let Us Study Georgian Folk Songs (Gurian Songs)* and *Teach Yourself Georgian Folk Songs - Megrelian Songs*, both published in 2004 by the International Centre for Georgian Folk Song. The performers are highly trained musicians and experts in the music of their regions. [1] The performers were recorded together as an ensemble in close proximity in a single room in a studio [15]. There was one microphone for each part and one microphone for the room. This system of recordings was intended as an aid for singers learning their parts on Georgian songs, a setting in which it is beneficial to hear all three parts, but one above the other two. All songs are in three parts except during short intervals of overlap between antiphonal ensembles or solo and trio. The ensemble for each song is one high tenor (*pirveli* in Georgian), one middle tenor (*meore*) and one bass (*bani*), with a few exceptions. The song "Khasanbegura" has two antiphonal ensembles, one a trio and the other consisting of one high tenor and a choir of middle tenors and basses. The three songs "Maq'ruli", "Orira", and "Shvidk'atsa" also have two ensembles, one a trio and the other consisting of two tenors and a bass choir. Two songs, "Indi-Mindi" and "P'at'ara Saq'varelo", have a fourth part, a brief solo bass, that we have ignored for purposes of f_0 estimation. All but the two songs "Sabodisho" and "Mi Re Sotsodali" are *a cappella*. The accompaniment in these songs is a *chonguri*, a picked Georgian lute.

1.2 f_0 Estimation

Research on monophonic f_0 estimation has developed over several decades. The work of [16] introduced the use of the "cepstrum" in f_0 estimation, which exploits the fact that harmonics are regularly spaced in the frequency domain. Building on the work of [17], which exploited the same fact using the technique of spectral compression, the work of [18] introduced the technique of subharmonic summation to model the ability of the human ear to detect a weak fundamental pitch and used numerical methods to reduce the susceptibility of the estimate to noise [19]. The work of [20] introduced an algorithm for f_0 estimation and voiced-unvoiced classification which used the autocorrelation function of the signal to generate candidate pitches and dynamic programming to generate a sequence of pitches with few large jumps in frequency and few isolated voiced or unvoiced points. The well-known Praat software uses this algorithm [21]. The work of [22] introduced the well-known Yin algorithm, which used the sum of squared differences between the signal and lagged signals instead of the autocorrelation function. The work of [23] introduced the use of the fast-lifting wavelet transform for f_0 estimation. The work of [24] introduced the use of a neural network for f_0 estimation with the CREPE algorithm.

We use these six algorithms – Noll [16], Hermes [18], Boersma [20], Yin, Maddox [23], and CREPE – as a "panel of experts" to estimate f_0 in these recordings. [2] One may anticipate that algorithms based on autocorrelation or deep learning are strictly better than those based on Fourier or wavelet analysis. However, we find that for the high-quality non-monophonic recordings of this dataset, each algorithm produces correct estimates that tend to persist through the duration of a musical note or longer, and the algorithms err under different conditions. The errors we observe are mainly pitches in the wrong octave and pitches on the wrong part, including pitch estimates at times when the current part is silent or is a fricative consonant. Despite the presence of such errors, on harmonic content (judged subjectively as vowels sung by a healthy voice) at least one of the algorithms typically estimates f_0 correctly.

The paper is arranged as follows. In Section 2 we give details about the data and describe the process and workflow. [3] In Section 3 we present visualizations that com-

[1] The performers on the Gurian songs include Guri Sikharulidze, Tristan Sikharulidze, Otar Berdzenishvili, Anzor Erkomaishvili, Gedevan Mzhavanadze, Kote Papava, and Levan Goliadze. The performers on the Megrelian songs include members of the Odoia Choir directed by Polikarpe Khubulava.

[2] pYin did not improve over Yin in initial testing.

[3] The code resides in a public Github respository [25] The raw data and annotations are available to researchers upon request.

pare the performance of the six estimation algorithms on the dataset and that support the use of median absolute deviation as a measure of confidence in the estimates. We also introduce two types visualization that shed some light on musical content of the collection. Finally we present an illustration of the web interface of the dataset. [4]

2. PROCESS AND WORKFLOW

2.1 Estimates and alignment

The first step in our process is to apply six algorithms for monophonic f_0 estimation to each recording. The outputs of these algorithms are shifted in time, typically by a few hundredths of a second. We align the six estimates in time by fixing the estimates of one arbitrarily-chosen algorithm (Boersma) and shifting each of the others by multiples of $0.01s$. We choose the shift that minimizes $d(\delta t)$, the time-lagged ℓ_2-distance between the two time series at lag δt. We limit our search to the range $[-0.1, 0.1]$ because the vibrato of the human voice is $5-7Hz$, which is the Nyquist frequency of a sample rate of $10 - 14Hz$ [27].

Figure 1 shows a plot of $d(\delta t)$ for one song in the collection. In most cases δt_0 is very close to 0.

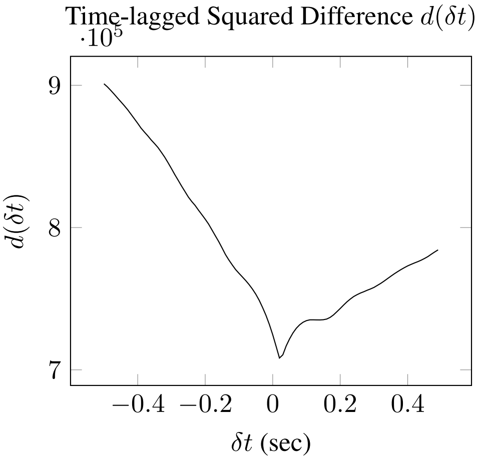

Figure 1. The squared difference between Boersma at time t and another f_0 estimator at time $t + \delta t$ for the middle part of *Ak'a Si Rekisho* from the Megrelian collection.

2.2 Note estimation

The next step in our process is a heuristic method for identifying notes in the target estimate. We chose a method that is simple to code and produces note values that serve well as a visual aid for the analyst, as opposed to a state-of-the-art method such as [28]. For purposes of post-processing non-note sections longer than $0.2s$ are considered unvoiced and the rest are considered voiced consonants.

The idea behind the heuristic is that a note transition is typically a monotone change in pitches between a local peak and a local trough in the singer's vibrato. The heuristic traverses the time series forward and identifies a note transitions as a change of 7% from the most recent local peak or trough. The pitch value of the note estimate is the average of the target estimate pitches over the duration of the note. The heuristic sets a minimum note length of $0.07s$. The heuristic also identifies as *transitions* monotone sequences of pitches between notes and increasing sequences of pitches before notes. Anything else is a non-note. We tuned the parameters during development. The 7% note transition threshold typically errs on the side of false negative note transitions.

2.3 Voiced-unvoiced detection

In parallel to the two steps just described, a heuristic "voiced/unvoiced" method removes non-salient voice parts that are present when the salient voice part is silent. This heuristic is a simple threshold based on the histogram of amplitude and the technique of weighted zero crossings. The recording is divided into overlapping frames of length $0.015s$ spaced $0.01s$ apart. For each frame the root mean squared of amplitude is divided by the zero crossing rate. A histogram of the ratio is created and smoothed using a Hanning window. The smoothed histogram typically has a local maximum near 0 that is due to sections of silence in the recording. There are typically one or two other local maxima near 0 that correspond to silence in the salient part. The "voiced" threshold is characterized by a high local maximum preceded by a low local minimum.

We tuned the parameters to result typically in some false positive sections of the recording that require manual correction but only isolated false negatives that can be corrected automatically. Figure 2 shows the smoothed histogram for one part of one song.

In Georgian music there is often a *decrescendo* in each voice at the end of a phrase. The threshold produced by the algorithm tends to correctly classify the ends of phrases but also tends to generate false *voiced* classifications during the subsequent silence. To combat this problem we added an automatic post-processing step which reclassifies short *voiced* sections within *unvoiced* sections.

We hand-tuned each of the heuristics mentioned above on a sample of varied musical sections, with the goal of minimizing the use of the interactive tool.

2.4 Interactive tool

The analyst uses a graphical interface to construct a target f_0 estimate for each part by stitching together the best estimates for different sections of the part. In sections of a part where no estimate is correct the analyst can specify a pitch range. The interface presents a window containing plots of the time series of f_0 estimates for one part of one song. The analyst can switch between parts and can turn the visibility of each time series on and off. The time series include the f_0 estimates of the six algorithms before and after application of the voiced-unvoiced algorithm. The window also includes the *target estimate*, which is initially equal to the CREPE time series, and note estimates for the target estimate. Selection tools and buttons enable the main functionalities of the tool, which are as follows:

[4] The web interface for the dataset makes available the f_0 estimates of the three vocal parts of each song [26].

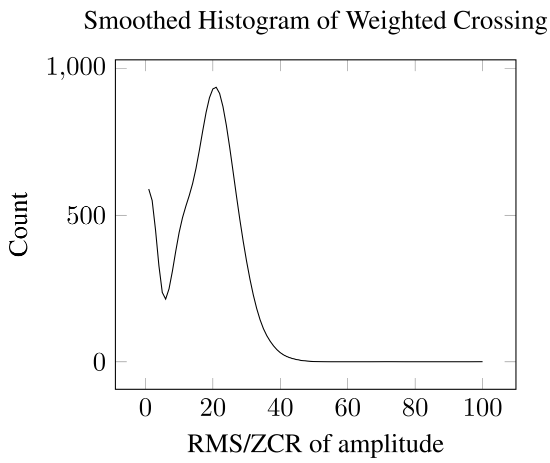

Figure 2. Smoothed histogram of weighted crossing for the top part of *Ak'a Si Rekisho* from the Megrelian collection

- to delete a section of the target estimate, i.e., set $f_0 = 0$ at selected times
- to set a section of the target estimate equal to the estimate of a selected algorithm
- to override a section of the target estimate with a pitch range

After using the tool on one part of a song, the analyst post-processes and saves the target estimate.

Figure 3 shows a screenshot of the tool in operation.

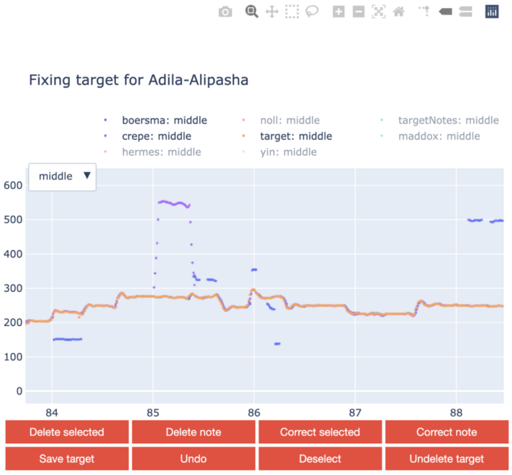

Figure 3. The interactive tool in operation at $84s$ in the middle part of *Adila-Alipasha* from the Gurian collection.

The last step of the process is to rerun the f_0 algorithms Noll, Hermes, Boersma, and Yin, constrained to be within a fixed percentage of the target estimate or within the pitch range. For CREPE and Maddox we use implementations that do not allow pitch constraints as input. If an algorithm is unable to find an estimate within the pitch range, its estimate is dropped.

2.5 Automatic Post-processing

The post-processing algorithm makes the following changes to the analyst's final choice of target estimate:

1. Revert isolated $(0.01 - 0.02s)$ unvoiced pitches to the estimates that were made before the application of the voiced-unvoiced algorithm.
2. Set unvoiced non-note sections (those shorter than $0.2s$) of the target estimate to 0.
3. Set voiced non-note sections of the target estimate to -10 to indicate *undecided*.

2.6 Manual Post-processing

The process described above generally results in a median absolute deviation of the estimates that is less than 1% of the median of estimates. (See Figure 6.) There are exceptions, sections of harmonic content where the estimates disagree. In some cases, typically sustained, well-tuned chords, it is clearly audible that two or more estimates are incorrect and skew the result. In these cases we have manually set a narrow pitch range in the interest of accuracy in the annotations.

3. RESULTS

In this section we discuss qualitatively the effectiveness of our method of combining multiple f_0 estimates. We provide descriptive statistics about the f_0 estimates that we include in the dataset of Gurian and Megrelian songs. First, we assess the estimates produced by each of the six algorithms on this dataset by comparing the initial estimate of each algorithm with the final estimate. Second, we assess the variability of the estimates as a measure of confidence in the final estimate. Third, we provide visualizations of the pitches and intervals of one song, *Gepshvat Ghvini*. Finally, we describe the web interface for the dataset and illustrate it with an example.

3.1 Discussion of the method

In most situations our process of combining multiple f_0 estimates "works" in the sense of producing near agreement among estimates where the fundamental frequency is unambiguous to the human ear. There is one noticeable failure mode. On non-monophonic input Hermes and Yin may fail to produce an estimate within given upper and lower frequency limits, and Boersma may label the sample *unvoiced*. In particular, Hermes, Yin, and Boersma tended to produce no estimate or incorrect estimates on prolonged, well-tuned chords. This is not a noticeable problem for Noll. We do not have a theoretical explanation, but we have found cases where the normalized difference function of Yin either does not have a minimum on the given interval or the location of minimum is unstable. On a small fraction of samples our process produced only two estimates or widely varying estimates.

3.2 Assessing each f_0 estimate

In this subsection we measure the accuracy of each monophonic f_0-estimation algorithm relative to the final f_0 estimate, separately on each collection. We consider the f_0 estimate of each algorithm, after applying alignment and the voiced-unvoiced algorithm, limited to samples that the note estimation algorithm labels as notes. Figures 4 and 5 show one smoothed histogram for each algorithm; it is a histogram of the ratio of the algorithm's f_0 estimate to the final f_0 estimate. By this measure Boersma is the most accurate when it is in the right octave (near 1.0), but Crepe and Noll are in the right octave more often.

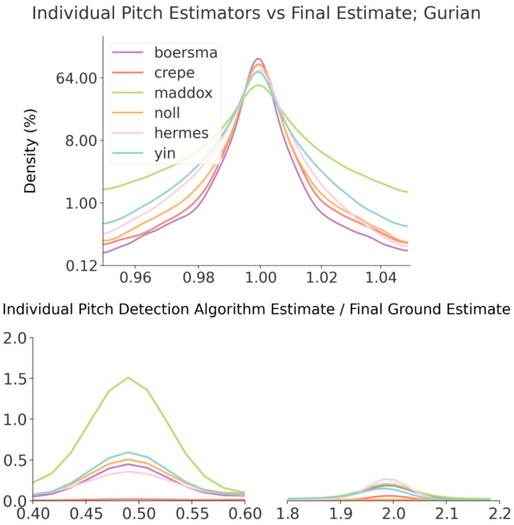

Figure 4. Smoothed histogram of ratio of f_0 estimates, algorithm/final, Gurian collection. Percentage of samples shown: boersma: 85.7%, crepe: 94.7%, maddox: 56.9%, noll: 85.9%, hermes: 86.0%, yin: 81.7%.

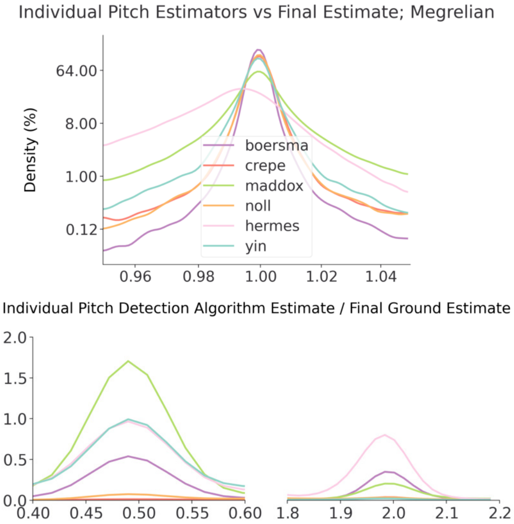

Figure 5. Smoothed histogram of ratio of f_0 estimates, algorithm/final, Megrelian collection. Percentage of samples shown: boersma: 84.3%, crepe: 96.3%, maddox: 63.1%, noll: 91.8%, hermes: 63.0%, yin: 79.4%.

3.3 Variability of the f_0 estimates

At each time step the final f_0 estimate is the median of the estimates of several algorithms. In this subsection we measure the variability of the estimates around the median. This measurement lends credibility to our method by showing that when there is a note (according to the note estimation algorithm) the variability is small (the estimates agree) and when there is not a note the variability is large (the estimates disagree). For our measure of variability we choose the median absolute deviation (MAD) around the median. This measure is robust to outliers, which means that when there is agreement among at least three estimates the MAD will be small. This choice corresponds to our intuition that during a vowel there will be general agreement among estimates and during a consonant there will be general disagreement. Figure 6 shows two graphs of smoothed histograms of MAD (in Hz), one for *note* samples and one for *non-note* samples. The graphs show 97.7% of Gurian note samples and 96.1% percent of Megrelian note samples. They show 70.9% percent of non-note Gurian samples and 74.7% of non-note Megrelian samples.

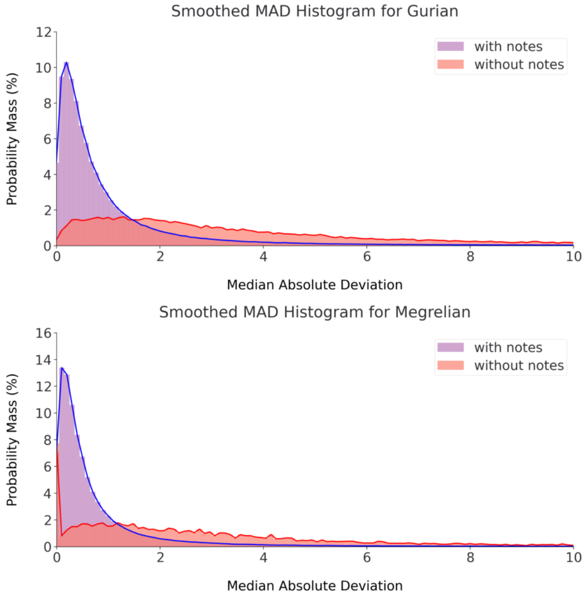

Figure 6. Smoothed histograms of MAD of each estimate, in Hz, *note* vs. *non-note* samples. Top: Gurian collection. Bottom: Megrelian collection.

3.4 Musical Observations

In this subsection we provide visualizations of the pitches and intervals in one Megrelian song, *Gepshvat Ghvini*. The purpose of these visualizations is to illustrate the potential for visualizations of digital data to enable and inform research in ethmomusicology.

We return to the question mentioned in the introduction of what intervals make up the Georgian scale, and to what extent singers deviate from the scale to produce desired harmonies. Figure 7 shows a histogram of the pitches of the three voices for *Gepshvat Ghvini*. The variation in the peaks of the histograms for the three voices between 210Hz and 220Hz possibly indicates deviation from a fixed scale. Further investigation is needed.

Figure 8 shows a two-dimensional histogram of the chords of *Gepshvat Ghvini*. The x-axis is the ratio of pitches between the middle and bass parts. The y-axis is the ratio of pitches between the top and middle parts. The histogram is shown as a heatmap. The dark patch roughly at coordinates $(1.22, 1.22)$ shows that triads with a perfect fifth are common, and that these triads cluster around those with a "neutral" third between the minor and major third.

To exclude pitches in transition between notes we have limited the data in both figures to pitches at least three time steps away from the first and last pitch of each note, as designated by the note estimation algorithm. In the two-dimensional histogram we have limited the data to samples where we have estimates for all three parts.

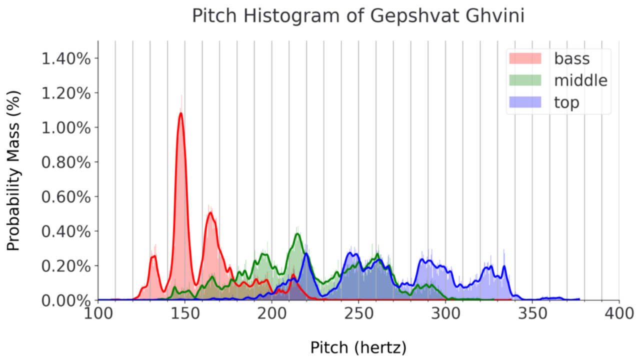

Figure 7. Histogram of pitches in *Gepshvat Ghvini*.

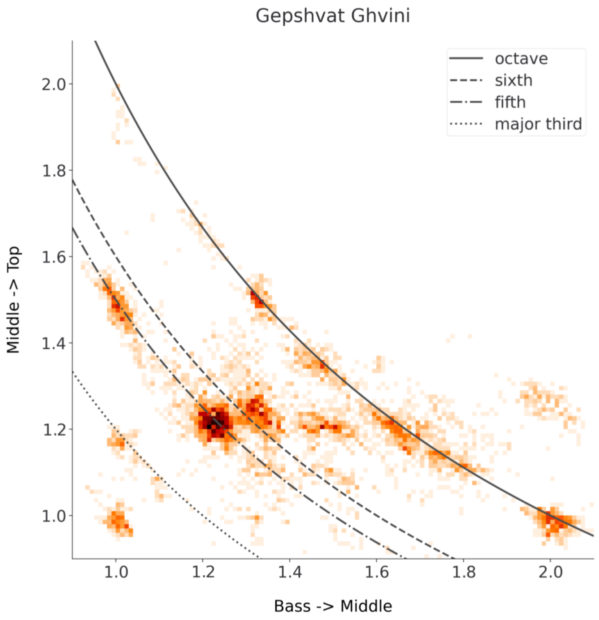

Figure 8. Histogram of chords in *Gepshvat Ghvini*. The x-axis is the ratio of middle to bass pitch. The y-axis is the ratio of top to middle pitch.

3.5 Web interface

In this subsection we describe the web interface for our dataset and illustrate the interface with an example. The web interface serves as a visual tool for the analyst to assess the accuracy of the results, as well as for ethnomusicologists to analyze of songs and for singers to learn the songs. The interface shows a scrolling graph of pitches that is synchronized to an audio player. Figure 9 shows a snapshot of the interface for the Megrelian song *Gepshvat Ghvini*. The graph optionally displays any of the median estimates of the three parts and the median absolute deviations (MAD) of the estimates.

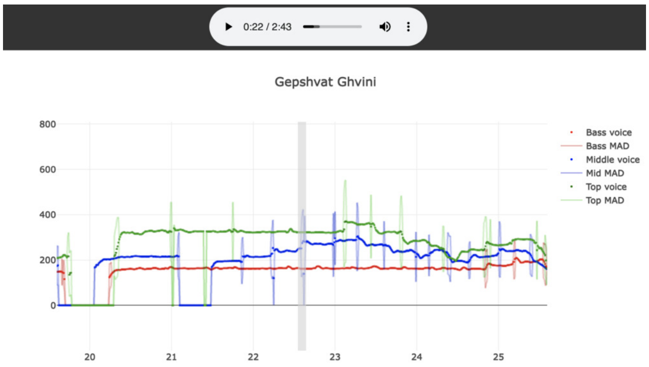

Figure 9. Web interface: *Gepshvat Ghvini*. The shaded line shows the current time. Each solid line shows the median frequency estimate for one part; the faint line of the same color shows the median absolute deviation.

4. CONCLUSION

We have presented a process and workflow for multi-f_0 annotation of a dataset of songs from the Republic of Georgia in which each song is represented by four recordings from different microphones. The annotations in our dataset represent the average of f_0 from different algorithms and are accompanied by the median absolute deviation of the estimates, which we have shown is a reasonable measure of confidence in the estimate. We hope our dataset and web interface are of interest to ethnomusicologists as audiovisual aids for analysis.

We plan to apply our process to collections of recordings from two other regions of Georgia which have been produced recently [29, 30]. We plan to use the resulting labeled datasets to develop algorithms for multi-f_0 estimation using the recordings of the mixed voices from the room microphones in our collections. We plan to develop the web interface further as a learning aid for singers by adding new features, including lyrics as subtitles.

5. ACKNOWLEDGEMENTS

The authors would like to thank the International Centre for Georgian Folk Song and the International Research Center for Traditional Polyphony, for permission to share the audio data with researchers.

6. REFERENCES

[1] R. Bittner, J. Salamon, M. Tierney, M. Mauch, C. Cannam, and J. Bello, “MedleyDb: A Multitrack Dataset for Annotation-Intensive MIR Research,” in *Proceedings of the 15th International Society for Music Information Retrieval Conference*, Taipei, Taiwan, Oct. 2014.

[2] S. Rosenzweig, H. Cuesta, C. Weiß, F. Scherbaum, E. Gómez, and M. Müller, “Dagstuhl ChoirSet: A Multitrack Dataset for MIR Research on Choral Singing,” *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, pp. 98–110, Jul. 2020, number: 1 Publisher: Ubiquity Press. [Online]. Available: http://transactions.ismir.net/articles/10.5334/tismir.48/

[3] S. Rosenzweig, F. Scherbaum, D. Shugliashvili, V. Arifi-Müller, and M. Müller, “Erkomaishvili Dataset: A Curated Corpus of Traditional Georgian Vocal Music for Computational Musicology,” *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, pp. 31–41, Apr. 2020, number: 1 Publisher: Ubiquity Press. [Online]. Available: http://transactions.ismir.net/articles/10.5334/tismir.44/

[4] F. Scherbaum, N. Mzhavanadze, S. Rosenzweig, and M. Müller, “Multi-Media Recordings Of Traditional Georgian Vocal Music For Computational Analysis,” in *Proceedings of the 9th International Workshop on Folk Music Analysis*, Birmingham, Jul. 2019.

[5] H. Cuesta, B. McFee, and E. Gómez, “Multiple F0 Estimation in Vocal Ensembles using Convolutional Neural Networks,” in *Proceedings of the 21st International Society for Music Information Retrieval Conference*, Montréal, Canada, 2020, arXiv: 2009.04172. [Online]. Available: http://arxiv.org/abs/2009.04172

[6] M. Mauch, C. Cannam, R. M. Bittner, G. Fazekas, J. Salamon, J. Dai, J. Bello, and S. Dixon, “Computer-aided Melody Note Transcription Using the Tony Software: Accuracy and Efficiency,” in *Conference Proceedings of the First International Conference on Technologies for Music Notation and Representation*, 2015.

[7] ——, “Tony: a tool for melody transcription, https://code.soundsoftware.ac.uk/projects/tony,” 2015. [Online]. Available: https://code.soundsoftware.ac.uk/projects/tony

[8] M. Müller, S. Rosenzweig, J. Driedger, and F. Scherbaum, “Interactive Fundamental Frequency Estimation with Applications to Ethnomusicological Research,” in *Proceedings of the Audio Engineering Society Conference: 2017 AES International Conference on Semantic Audio*. Audio Engineering Society, Jun. 2017. [Online]. Available: https://www.aes.org/e-lib/browse.cfm?elib=18777

[9] F. Scherbaum, N. Mzhavanadze, S. Arom, S. Rosenzweig, and M. Müller, “Tonal Organization of the Erkomaishvili Dataset: Pitches, Scales, Melodies and Harmonies,” *Computational Analysis Of Traditional Georgian Vocal Music*, 2020, publisher: Universitätsverlag Potsdam. [Online]. Available: https://publishup.uni-potsdam.de/frontdoor/index/index/docId/47614

[10] A. Erkomaishvili, *Georgian Folk Music. Guria. From Artem Erkomaishvili’s Collection.* Tbilisi, Georgia: International Centre for Georgian Folk Song, 2005.

[11] S. Gelzer, “Testing a Scale Theory for Georgian Folk Music,” in *The First International Symposium on Traditional Polyphony, Proceedings*. International Research Center for Traditional Polyphony, Sep. 2002, pp. 194–200. [Online]. Available: https://drive.google.com/file/d/1t1wNS_d2mBnPzeZnqSOe8QUymowiriwU/view?usp=drive_open&usp=embed_facebook

[12] M. Erkvanidze, “On Georgian Scale System,” in *The First International Symposium on Traditional Polyphony, Proceedings*. International Research Center for Traditional Polyphony, Sep. 2002, pp. 178–185. [Online]. Available: https://drive.google.com/file/d/1MT2fdZwhrkMitpDjDYZBwTtxS_jl1DT0/view?usp=drive_open&usp=embed_facebook

[13] Z. Tsereteli and L. Veshapidze, “On the Georgian Traditional Scale,” in *The Seventh International Symposium on Traditional Polyphony, Proceedings*. Tbilisi, Georgia: International Research Center for Traditional Polyphony, Sep. 2014, pp. 288–295. [Online]. Available: https://drive.google.com/file/d/1iVzCX1jAXnMnv_rJWNYEDdtOPTARqUzj/view?usp=drive_open&usp=embed_facebook

[14] D. Shugliashvili, *Georgian Church Hymns, Shemokmedi School.* Georgian Chanting Foundation, 2014.

[15] N. Razmadze, “Private communication,” Mar. 2022.

[16] A. M. Noll, “Short-Time Spectrum and "Cepstrum" Techniques for Vocal-Pitch Detection,” *The Journal of the Acoustical Society of America*, vol. 36, no. 296, 1964.

[17] ——, “Pitch determination of human speech by the harmonic product spectrum, the harmonic sum spectrum, and a maximum likelihood estimate,” in *Proceedings of the Symposium on Computer Processing in Communication*, vol. 19. New York, NY, USA: Microwave Institute, University of Brooklyn, 1970, pp. 779–797.

[18] D. Hermes, “Measurement of pitch by subharmonic summation,” *The Journal of the Acoustical Society of America*, vol. 83, no. 257, 1988.

[19] E. Terhardt, G. Stoll, and M. Seewann, “Algorithm for extraction of pitch and pitch salience from complex tonal signals,” *The Journal of the Acoustical Society of America*, vol. 71, pp. 679–688, 1982.

[20] P. Boersma, “Accurate short-term analysis of the fundamental frequency and the harmonics-to-noise ratio of a sampled sound,” in *IFA Proceedings 17*, 1993, pp. 97–110.

[21] P. Boersma and D. Weenink, “Praat, a system for doing phonetics by computer,” *Glot International*, vol. 5, no. 9/10, pp. 341–345, 2001.

[22] A. de Cheveigné and H. Kawahara, “YIN, a fundamental frequency estimator for speech and music,” *The Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1917–1930, Apr. 2002, publisher: Acoustical Society of America. [Online]. Available: https://asa.scitation.org/doi/10.1121/1.1458024

[23] E. Larson and R. K. Maddox, “Real-Time Time-Domain Pitch Tracking Using Wavelets,” in *Proceedings of the University of Illinois at Urbana Champaign Research Experience for Undergraduates Program*, 2005.

[24] J. W. Kim, J. Salamon, P. Li, and J. Bello, “Crepe: A Convolutional Representation for Pitch Estimation,” *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2018.

[25] D. Gillman, U. Goyat, and A. Kutlay, “Teach Yourself Georgian Folk Songs Dataset: Code,” 2021. [Online]. Available: https://github.com/New-College-of-Florida/Voice

[26] ——, “Teach Yourself Georgian Folk Songs Dataset: Website,” 2021. [Online]. Available: http://131.247.152.67/georgian

[27] M. Vetterli, J. Kovačević, and V. K. Goyal, *Foundations of Signal Processing*. Cambridge University Press, Sep. 2014. [Online]. Available: https://www.cambridge.org/highereducation/books/foundations-of-signal-processing/DCC08E20D354F34E084FC11862E18F18

[28] R. Nishikimi, E. Nakamura, K. Itoyama, and K. Yoshii, “MUSICAL NOTE ESTIMATION FOR F0 TRAJECTORIES OF SINGING VOICES BASED ON A BAYESIAN SEMI-BEAT-SYNCHRONOUS HMM,” *New York City*, p. 7, 2016.

[29] M. Khardziani and N. Razmadze, “Acharan Folk Songs – Collection Of Sheet Music With CD For Self-Study – International Research Center for Traditional Polyphony,” 2020. [Online]. Available: http://polyphony.ge/en/acharan-folk-songs-collection-of-sheet-music-with-cd-for-self-study/

[30] R. Tsurtsumia and N. Razmadze, “Svan Folk Songs,” May 2020, section: Books. [Online]. Available: https://chanting.ge/en/svan-folk-songs/