Audio Engineering Society

Conference Paper

Presented at the Conference on
Semantic Audio
2017 June 22 – 24, Erlangen, Germany

Interactive Fundamental Frequency Estimation with Applications to Ethnomusicological Research

Meinard Müller[1], Sebastian Rosenzweig[1], Jonathan Driedger[1], and Frank Scherbaum[2]

[1] *International Audio Laboratories Erlangen, Germany*
[2] *Institut für Erd- und Umweltwissenschaften, Universität Potsdam, Germany*

Correspondence should be addressed to Meinard Müller (`meinard.mueller@audiolabs-erlangen.de`)

ABSTRACT

The analysis of recorded audio sources has become increasingly important in ethnomusicological research. Such audio material may contain important cues on performance practice, information that is often lost in manually generated symbolic music transcriptions. As an application scenario, we consider in this paper a musically relevant audio collection that consists of three-voice polyphonic Georgian chants. As one main contribution, we introduce an interactive graphical user interface that provides various visual and acoustic control mechanisms for estimating fundamental frequency (F0) trajectories from complex sound mixtures. We then apply this interface for determining F0 trajectories of sung pitches from the Georgian chant recordings and indicate how such F0 annotations can be used as basis for addressing important questions in Georgian music research.

1 Introduction

Ethnomusicological research is typically conducted on the basis of notated music material, which is obtained by transcribing recorded tunes into symbolic, score-based music representations. These transcriptions are often idealized and tend to represent the presumed intention of the singer rather than the actual performance. As a result, performance aspects enclosed in the recorded audio material may be lost in symbolic music representations [1, 2]. Therefore, when researching performance practice, the analysis of recorded music material seems inevitable. However, opposed to symbolic representations, musical parameters such as pitches, note onsets, or note durations are not given explicitly in an audio representation, which basically encodes the acoustic waveform of a performance. The manual extraction of musical parameters directly from a music recording is cumbersome and time consuming, thus asking for computer-based methods to assist an ethnomusicologist in accessing and understanding the audio material. In this paper, we investigate the applicability of automated methods to analyze a music collection that consists of audio recordings of traditional Georgian chants.

Georgia has a long and rich music history. In particular, Georgian polyphonic singing has been acknowledged as an intangible cultural heritage, which has “regained a place of prominence in the hearts and minds of the public and in the life of the Church” [3]. To preserve this cultural treasure, the “Commission for Chant Preservation” began in the 1860s the process of notating Georgian chants, which had been passed down orally for many generations. This has resulted in thousands of transcriptions collected between 1880–1920. Due to changing social and political conditions, however, the tradition of how to perform these chants has largely been lost, and ethnomusicologists have started to research on traditional performance practice [3]. In this context, a collection of music recordings performed by Artem Erkomaishvili (1887–1967)—one of the last representatives of the master chanters (“sruligalobelni”) of Georgian music—has become of great importance. Recorded at the Tbilisi State Conservatory in 1966, the aging Erkomaishvili was asked to perform three-voice chants by successively singing the individual voices. After recording the lead voice, one tape recorder was used to playback this first voice while a second tape recorder was used to synchronously record the middle voice. Similarly, playing back the first and second voice, the bass voice was recorded, see Figure 1a. In this way, Erkomaishvil was able to accompany and embellish his own recordings, yielding a genuine source of original Georgian musical thinking [3]. The resulting collection of ca. 100 audio recordings,[1] which comprises various types of chants including hymns for Easter, Christmas, or wedding ceremonies, is of great importance for ethnomusicological research.

To make this audio collection better accessible for musicological research, one important task is to estimate the fundamental frequency (F0) trajectories of the sung pitches from the recordings using automated methods. While this is feasible with standard procedures in the case of monophonic music, the problem becomes much harder in the case of polyphonic music. As one main contribution of this paper, we introduce a graphical user interface (GUI) for semi-automatic estimation of F0 trajectories. The GUI allows the user to specify temporal-spectral constraint regions that guide the estimation process. Furthermore, the GUI provides visual and acoustic feedback mechanisms that can be used to control and refine the estimated results in an interactive fashion. As an example scenario of musical relevance, we then apply this interface for extracting the F0 trajectories of the sung pitches from the three-voice Georgian chants recordings. Finally, we indicate how these F0 annotations may help to answer musicological questions within Georgian music research. In the remainder of this paper, we first review the F0-estimation procedure used in this work (Section 2), then introduce the GUI (Section 3), and finally consider as an application the Georgian chant scenario (Section 4). Further related work is discussed in the respective sections.

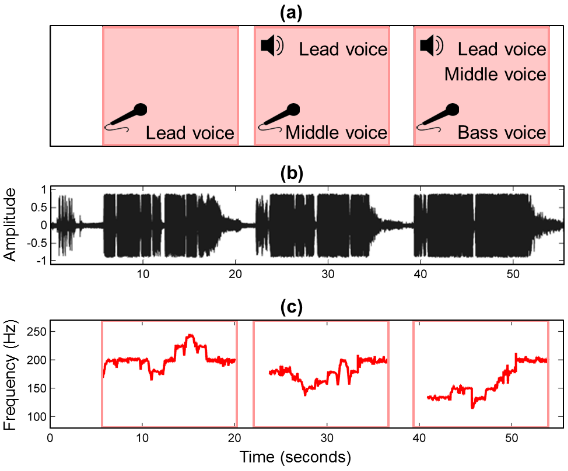


Fig. 1: Recording of the Georgian chant “Da Sulisatsa” sung by Artem Erkomaishvili. **(a)** Three-stage recording process. **(b)** Waveform. **(c)** Fundamental frequency trajectories for the lead voice, the middle voice, and the bass voice. The pink rectangles indicate the recording’s structure based on the three-stage recording process.

2 Fundamental Frequency Estimation

In this section, we give some background information on melody extraction and F0 estimation (Section 2.1) and then summarize the F0-estimation procedure used in this study (Section 2.2).

2.1 Background

As mentioned in the introduction, audio recordings (given as waveforms) are complex in the sense that musical parameters such as pitches, note onsets, or note durations are not given explicitly. Furthermore, real-world sounds are far from being simple pure tones with well-defined frequencies. Playing a single note on an instrument may result in a complex sound that contains a mixture of different frequencies changing over time. Intuitively, such a musical tone can be described as a superposition of pure tones or sinusoids, each with its own frequency of vibration, amplitude, and phase. A

[1] The audio collection is hosted at the Folklore Department, Tbilisi State Conservatory. It has been made available at `http://www.alazani.ge/old-archives-Artem-Erkomaishvilis-Sagaloblebi-folk-songs-ans59.html`

partial is any of the sinusoids by which a musical tone is described. The frequency of the lowest partial present is called the *fundamental frequency* (abbreviated as F0) of the sound. The *pitch* of a musical tone is usually determined by the fundamental frequency, which is the one created by vibration over the full length of a string or air column of an instrument. For further details, we refer to [4, Chapter 1].

When given an audio recording, one central task in music processing is referred to as *melody extraction*. In general terms, a *melody* (or more general a melodic voice) may be defined as a linear succession of musical tones expressing a particular musical idea. Because of the special arrangement of tones, a melody is perceived as a coherent entity. When being performed by a singer or played on an instrument, the melody corresponds to a sequence of frequency values rather than notes. Also, as opposed to a notated symbolic representation, some of the notes may be smoothly connected (e.g., when singing a glissando). Furthermore, one may observe rather pronounced frequency modulations due to vibrato. Given an audio recording, melody extraction is often understood as the task of estimating the sequence of frequency values that correspond to the main melody [5, 6, 7]. In other words, rather than estimating a sequence of notes, the objective is to determine a sequence of frequency values that correspond to the notes' pitches. Such a frequency path over time, which may also capture continuous frequency glides and modulations, is referred to as a *frequency trajectory*. In particular, one is interested in the fundamental frequency values of a melody's notes. The resulting trajectory is also called the melody's F0-trajectory. For further details, we refer to [4, Chapter 8]. Furthermore, the article by Salamon et al. [7] gives a comprehensive overview of melody extraction techniques and their applications.

The estimation of the fundamental frequency of a quasiperiodic signal, termed *monopitch estimation*, is a long-studied problem with applications in speech processing. For a review of early contributions we refer to [8]. While monopitch estimation is now achievable with a reasonably high accuracy, the problem of *multipitch estimation* with the objective of estimating the fundamental frequencies of concurrent periodic sounds remains very challenging. This particularly holds for music signals, where concurrent notes stand in close harmonic relation. For extreme cases such as complex orchestral music where one has a high level of polyphony, multipitch estimation becomes intractable with today's methods. Reviews of recent approaches can be found in [9, 10].

2.2 F0-Estimation Procedure

When extracting dominant fundamental frequency information from a complex, possibly polyphonic music recording, most approaches typically proceed in two steps. In the first step, the audio recording is converted into some kind of time–frequency representation. Then, in the second step, the dominant frequency values are selected for each time position, where one typically introduces temporal continuity conditions and exploits additional knowledge on the expected frequency range. Following this basic approach, we now summarize such a procedure closely following the work by Salamon et al. [11]. Our procedure is illustrated by Figure 2, where the three-stage recording of the Georgian chant "Da Sulisatsa" sung by Artem Erkomaishvili serves as our running example (see also Figure 1).

In the first step, the waveform is converted into a time-frequency representation by applying a suitable *short-time Fourier transform* (STFT) [4, 12]. By considering the (squared) magnitude of the STFT, one obtains a spectrogram representation (see Figure 2a). When used for extracting fundamental frequency information, the spectrogram representation is typically enhanced to better account for acoustic characteristics that are of perceptual and musical relevance. First, motivated by the observation that spectral components can show extremely small values while still being relevant for a human listener, we apply a technique referred to as *logarithmic compression* [13] to balance out the difference between large and small values. Second, accounting for the fact that the human sensation of sound height is logarithmic in frequency, we convert the linear frequency axis of a given spectrogram into a logarithmically spaced frequency axis that reflects the logarithmic nature of musical pitches. The resulting representation is often referred to as *log-frequency spectrogram*. The third enhancement strategy is based on the observation that a sound event such as a musical tone is associated to a fundamental frequency along with its harmonic partials, which are (approximately) the integer multiples of the fundamental frequency. The multiple appearances of tonal time–frequency patterns can be exploited to enhance a spectrogram representation by jointly considering a frequency and its harmonics forming suitably

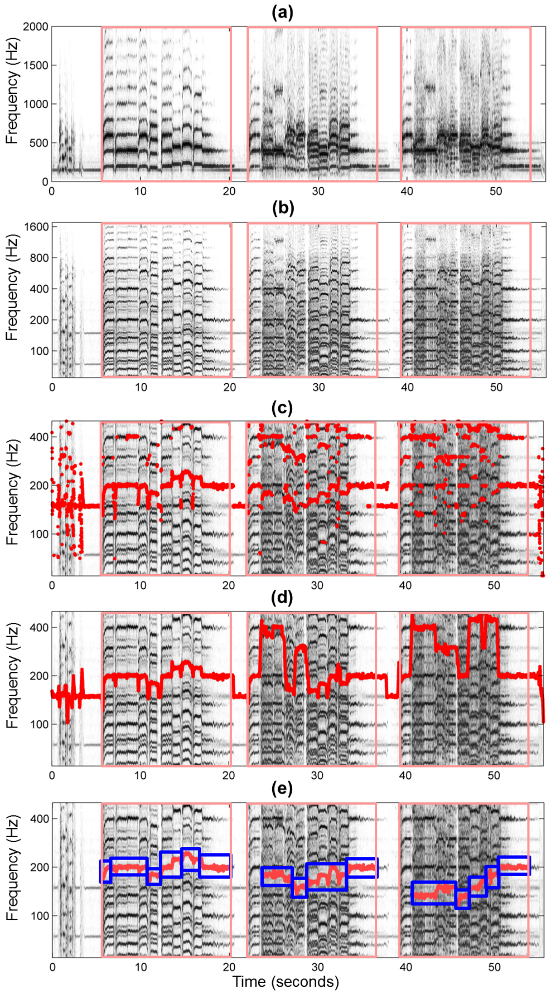


Fig. 2: Illustration of the F0 trajectory computation for the three-stage recording of Figure 1. **(a)** Spectrogram representation. The intensity is reflected by the shade of gray (the darker the more intense). **(b)** Salience representation (enhanced log-frequency spectrogram). **(c)** Frame-wise F0-trajectory (red line). **(d)** F0-trajectory with continuity constraints. **(e)** F0-trajectory with constrained regions (blue boxes).

weighted sums—a technique also called *harmonic summation*, see [5, 14, 11]. The resulting time-frequency representation is often referred to as *salience spectrogram*, since it makes the time-frequency coefficients that are likely to be part of a melody's F0-trajectory more salient (see Figure 2b). For further details, we refer to [4, Chapter 8] and [11].

In the second step, the goal is to determine relevant frequency information. Based on the assumption that the melody often correlates to the predominant F0-trajectory, a first strategy is to simply consider the frame-wise maximum of the computed salience representation (see Figure 2c). Such a simple frame-wise approach may lead to a number of temporal discontinuities and random jumps that occur due to confusions between the fundamental frequency and higher harmonics or lower ghost components introduced by the harmonic summation. To balance out the two conflicting conditions of temporal flexibility (to account for possible jumps between notes) and temporal continuity (to account for smoothness properties), one can use a procedure for constructing a frequency trajectory based on *dynamic programming* [15, 16] (see Figure 2d). Even though this may be a desirable property most of the time, discontinuities that are the result of abrupt note changes tend to be smoothed out. Furthermore, tracking errors still occur when there are several melodic lines or when there is no melody at all. Therefore, another strategy is to exploit additional musical knowledge about the melodic progression to support the F0-tracking process. For example, knowing the vocal range of the melodic voice, one may narrow down the search range of the expected F0-values. Or, having information about when the melody is actually present and when it is not, one can suppress the F0 estimation for non-melody frames. Additional knowledge as described above can be used to define *constraint regions* within the time–frequency plane. The F0-tracking is then performed only in these specified regions. Figure 2e shows an example, where such constraint regions are specified by the rectangular blue boxes. These boxes may be manually specified by a user (see also Section 3) or may be derived from available synchronized score information that underlies the given music recording [17].

3 Interactive F0 Estimation

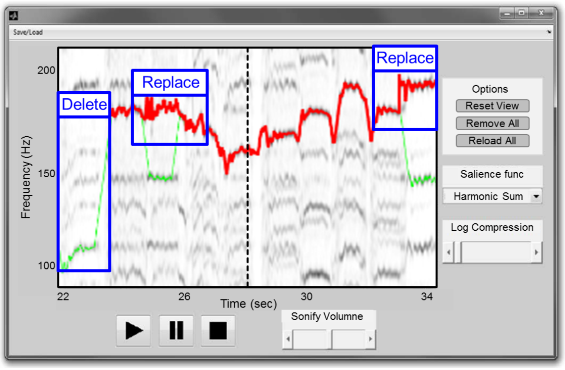
Due to acoustic and musical reasons, the automated extraction of fundamental frequencies is prone to errors—in particular for polyphonic music. Besides integrating strong musical assumptions and exploiting specific recording conditions, it is often necessary to integrate a user in the estimation process to validate and possibly correct the extraction results. In this section, we first discuss various software programs that can be used for the interactive computation of frequency trajectories (Section 3.1) and then present our GUI with its feedback mechanisms (Section 3.2).

3.1 Related Work

In the following, we give a small overview of commercial as well as academic software packages that are related to the extraction of fundamental frequency information. We start with a commercial audio processing software called *Melodyne*, which is a product by the company *Celemony*. This software allows the decomposition of a music recording into note-like audio events (referred to as *blobs*).[2] In the decomposition process, the software computes a fundamental frequency trajectory for each of these blobs. The decomposition itself can be influenced by changing the parameters that are used for the signal analysis or by providing prior information about the musical piece such as its key. Changing the settings also has an influence on the derived frequency trajectories. These trajectories, however, only serve for visualization purposes and cannot be exported.

Other examples for non-commercial tools that allow for an interactive derivation of frequency trajectories are *Tony* by Mauch et al. [18] or the interface introduced by Pant et al. [19]. After having analyzed a given music recording, these programs offer a choice of different frequency trajectories. A user can then select the trajectory that matches best the recording or intended application. One benefit of this approach is that the user only has to verify the correctness of a small number of computed trajectories, which makes this approach time efficient. However, especially when dealing with complex music recordings that are highly polyphonic, it may happen that none of the offered melody trajectories appropriately reflects the recording's actual melody or desired melodic voice.

Another software program that is very popular in academia is *Praat* [20]. This tool was particularly developed for the phonetic analysis of speech and also offers the possibility to estimate F0-trajectories. A user can influence the estimation process, for example by specifying the trajectory's expected frequency range. However, Praat does not seem to be suitable for the analysis of complex music recordings as the underlying F0-estimation procedure is particularly designed for monophonic recordings.



Salience func Harmonic Sum

Fig. 3: Graphical user interface for an interactive estimation of F0-trajectories.

3.2 Proposed GUI

To overcome some of the above mentioned issues, we now describe a graphical user interface (GUI) that allows a user to interactively correct frequency trajectories in a more local fashion. Figure 3 shows a screenshot of this GUI, which integrates the salience representation from Section 2.2 as its central visual element. On top of this representation, a previously specified frequency trajectory can be plotted. The GUI integrates the features of a standard audio player (see the buttons for starting, pausing, and stopping the playback of the loaded music recording at the bottom of the interface). When playing back the music recording, the respective time position is indicated by a vertical dashed playback bar running across the salience representation. This way, salient structures in the visual representation can be directly compared to the auditory cues in the recording. Additionally, the interface allows for playing back a sinusoidal sonification of the specified frequency trajectory (acoustically superimposed with the original audio recording). This way, simply by listening, the user can easily understand the accuracy of the current trajectory.

As another important feature, the GUI allows a user to correct a given frequency trajectory. To this end, the rough location of a frequency trajectory can be specified by means of a rectangular box (as indicated by the blue boxes). These boxes are used as constraint regions to recompute frequency trajectories within these regions, where previously specified trajectories within the corresponding time windows are replaced. To account for extremely fine-grained corrections, the user

[2] http://www.celemony.com/de/melodyne/what-is-melodyne

may even use an editing option for drawing the trajectory free hand. To further simplify the tracking process, the user can also visually enhance interesting structures in the salience representation by applying a logarithmic compression. Of course, it is possible to save the current state of the frequency estimation and correction process at any time and to resume the interactive process at a later stage.

Compared to other software packages, the proposed GUI may be more time-consuming, in particular when a user needs to estimate a melody trajectory from scratch by manually defining constraint regions. Especially for complex music recordings it is very likely that the number of constraint regions that are necessary to yield an appropriate frequency trajectory is high. On the other hand, our approach allows a user to generate melody trajectories of high quality, even for polyphonic recordings.

4 Application: Georgian Music Research

Despite its small size, Georgia is the home region of a large number of stylistically very diverse singing traditions which form an essential part of the cultural identity of this country and which are increasingly receiving attention of international music lovers, musicians, and scholars alike [21]. The distinctiveness of Georgian vocal polyphonic music in comparison to Western music is based on the abundant use of “dissonances” and on the fact that the music is not tuned to the 12-tone equal-tempered scale. While the non-tempered nature of traditional Georgian vocal music can be considered consensus among musicologists, the particular nature of the traditional Georgian tuning is an ongoing topic of intense and controversial discussions [22, 23, 24]. Based on the analysis of field recordings in Svaneti in 2015, for which larynx microphones with a nearly perfect voice separation during recording of the individual voices were used, Scherbaum [24] suggested that at least part of this “Georgian sound-scale controversy” might be due to differences between interval sizes in a melody (horizontal perspective) and interval sizes in a chord (vertical perspective).

Since Georgian vocal polyphony is primarily oral tradition music, historical recordings play a crucial role in trying to understand and possibly preserve the tuning systems of the past. In this context, the collection of ca. 100 audio recordings by Artem Erkomaishvili, which we described in Section 1, is of extraordinary importance since it provides a glimpse into the harmonic and melodic thinking of one of the last Georgian master chanters of modern Georgia.

To make this dataset accessible for musicological research, we applied our GUI to extract the F0-trajectories for these recordings. This task was done by a user, who had some musical background (an amateur musician), but had no specific training in signal processing or computer science. In a first step, the user determined the recordings’ structures based on the three-stage recording process (see Figure 1). Subsequently, the user determined the F0-trajectories for the lead, middle, and bass voices from the first, second, and third section, respectively. To this end, the salience visualization and sonification functionality helped the user to determine suitable constraint regions to guide the estimation process. Note that the time-frequency constraints helped to roughly locate the relevant information, while the detailed frequency estimation within the constraint regions was done automatically by the F0-extraction procedure as described in Section 2.2. All results, including the original recordings, figures of the salience representations, the estimated F0-trajectories, and the sonifications of these trajectories, have been made publicly available[3].

Because of the historical importance of Erkomaishvili’s recordings, the results of our F0 estimation may serve as a starting point for a whole set of subsequent analysis steps to address a number of ethnomusicological issues including the analysis of the historical tuning system, transcription-free documentation, harmonic analysis, and quantitative comparison of chants just to name a few. As an example, we compute the chord-based interval size distribution for the three-voice chants (vertical perspective) similar to the approach suggested by Scherbaum [24]. Using our running example from Figure 1, the procedure is illustrated by Figure 4. First, the estimated F0-trajectories of the lead, middle, and bass voice (see Figure 4a) are superimposed (see Figure 4b). Then, for each time position, the intervals (given in cents) between the F0-trajectories of the lead and middle voice, the lead and bass voice, as well as the middle and bass voice are computed (see Figure 4c). Finally, integrating the occurrences of the different intervals over time, we obtain

[3] `https://www.audiolabs-erlangen.de/resources/MIR/2017-GeorgianMusic-Erkomaishvili`

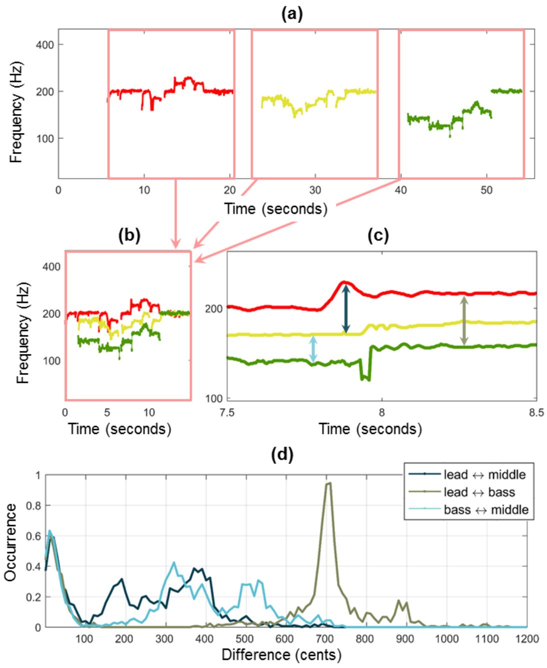


Fig. 4: **(a)** Estimated F0 trajectory for the running example (see Figure 1). **(b)** Superposition of the F0 tracks estimated for the three voices. **(c)** Illustration of the interval computation (using a zoomed section of (b)). **(d)** Interval distribution.

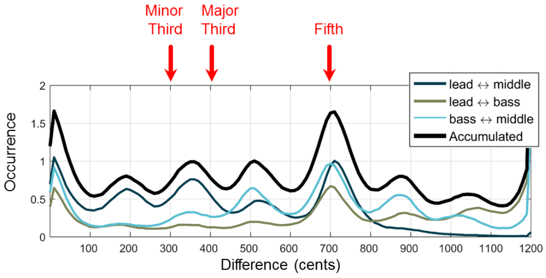


Fig. 5: Interval distributions obtained by considering all recordings by Erkomaishvili.

for each of the three cases an interval distribution (see Figure 4d).

In our experiments, we computed and averaged such distributions over the recordings by Erkomaishvili. The three resulting average distributions along with an accumulated distribution (considering all three cases jointly) are shown in Figure 5. Looking at these distributions, one can make some interesting observations. Disregarding the peaks close to the unison interval (0 cents) and octave interval (1200 cents), the most prominent peak occurs close to the fifth interval (702 cents in just intonation, 700 cents in the 12-tone equal-tempered scale). This reflects the fact that the fifth interval plays an important in Georgian chants and that this interval is sung with high intonation accuracy. Interestingly, there is another noticeable peak located at about 350 cents. From a Western music perspective, this is an usual interval since it lies between the minor third (315.6 cents in just intonation, 300 cents in the 12-tone equal-tempered scale) and the major third (400 cents in the 12-tone equal-tempered scale). The peak may be the result of the non-tempered nature of traditional Georgian vocal music [24]. From a Western music perspective, the role of the third interval in Georgian music seems ambiguous and, in combination with a fifth, evokes in the listener a sound that somehow lies between a minor and major chord. These observations agree with the results of the much more detailed study by Scherbaum [24], which was conducted on recent larynx-microphone field recordings. Our F0 annotations may facilitate similar studies on the “Georgian sound-scale controversy” based on Erkomaishvili’s historical recordings, thus adding a historical perspective on this important ethnomusicological issue.

5 Conclusions

In this paper, we introduced a GUI that allows a user to estimate F0-trajectories for complex, possibly polyphonic music recordings. As an illustrating application scenario, we estimated F0-trajectories for a data collection of historical importance consisting of roughly 100 three-voice Goergian chant recordings. All generated annotations along with visualizations and sonifications have been made publicly available. Finally, we indicated the potential of the annotations for musicological research. We hope that our contributions yield a starting point for further ethnomusicological studies within Georgian music research and beyond.

Acknowledgments

This work has been supported by the German Research Foundation (DFG MU 2686/6-1). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institut für Integrierte Schaltungen IIS. We thank Stefanie Kämmerle, Vlora

Arifi-Müller, and Stefan Balke for helping us with the annotations, the database setup, and the websites.

References

[1] Tzanetakis, G., "Computational ethnomusicology: a music information retrieval perspective," in *Music Technology meets Philosophy - From Digital Echos to Virtual Ethos: Joint Proceedings of the 40th International Computer Music Conference (ICMC) and the 11th Sound and Music Computing Conference (SMC)*, Michigan Publishing, Athens, Greece, 2014.

[2] Müller, M., Grosche, P., and Wiering, F., "Automated analysis of performance variations in folk song recordings," in *Proceedings of the International Conference on Multimedia Information Retrieval (MIR)*, pp. 247–256, Philadelphia, Pennsylvania, USA, 2010.

[3] Shugliashvili, D., "Introduction," in *Georgian Church Hymns, Shemokmedi School*, pp. 23–29, 2014.

[4] Müller, M., *Fundamentals of Music Processing*, Springer Verlag, 2015, ISBN 978-3-319-21944-8.

[5] Goto, M., "A Real-time Music-scene-description System: Predominant-F0 Estimation for Detecting Melody and Bass Lines in Real-world Audio Signals," *Speech Communication (ISCA Journal)*, 43(4), pp. 311–329, 2004.

[6] Poliner, G. E., Ellis, D. P., Ehmann, A. F., Gómez, E., Streich, S., and Ong, B., "Melody Transcription From Music Audio: Approaches and Evaluation," *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4), pp. 1247–1256, 2007.

[7] Salamon, J., Gómez, E., Ellis, D. P. W., and Richard, G., "Melody Extraction from Polyphonic Music Signals: Approaches, applications, and challenges," *IEEE Signal Processing Magazine*, 31(2), pp. 118–134, 2014, doi: 10.1109/MSP.2013.2271648.

[8] Hess, W., *Pitch Determination of Speech Signals*, Springer-Verlag, Berlin, 1983.

[9] de Cheveigne, A., "Multiple F0 estimation," in D. Wang and G. J. Brown, editors, *Computational Auditory Scene Analysis*, Wiley/IEEE Press, 2006.

[10] Klapuri, A. P. and Davy, M., editors, *Signal Processing Methods for Music Transcription*, Springer, New York, 2006, ISBN 0-387-30667-6.

[11] Salamon, J. and Gómez, E., "Melody Extraction from Polyphonic Music Signals using Pitch Contour Characteristics," *IEEE Transactions on Audio, Speech, and Language Processing*, 20(6), pp. 1759–1770, 2012.

[12] Gabor, D., "Theory of Communication," *Journal of the Institution of Electrical Engineers (IEE)*, 93(26), pp. 429–457, 1946.

[13] Klapuri, A. P., "Multipitch Analysis of Polyphonic Music and Speech Signals Using an Auditory Model," *IEEE Transactions on Audio, Speech, and Language Processing*, 16(2), pp. 255–266, 2008.

[14] Klapuri, A. P., "Multiple fundamental frequency estimation by summing harmonic amplitudes," in *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pp. 216–221, 2006.

[15] Rabiner, L. and Juang, B.-H., *Fundamentals of Speech Recognition*, Prentice Hall Signal Processing Series, 1993.

[16] Müller, M., *Information Retrieval for Music and Motion*, Springer Verlag, 2007, ISBN 3540740473.

[17] Driedger, J., Grohganz, H., Prätzlich, T., Ewert, S., and Müller, M., "Score-Informed Audio Decomposition and Applications," in *Proceedings of the ACM International Conference on Multimedia (ACM-MM)*, pp. 541–544, Barcelona, Spain, 2013.

[18] Mauch, M., Cannam, C., and Fazekas, G., "Efficient computer-aided pitch track and note estimation for scientific applications," 2014, extended abstract accepted at SEMPRE 2014 conference.

[19] Pant, S., Rao, V., and Rao, P., "A melody detection user interface for polyphonic music," in *National Conference on Communications (NCC)*, pp. 1–5, 2010.

[20] Boersma, P., "Praat, a system for doing phonetics by computer," *Glot International*, 5(9/10), pp. 341–345, 2001.

[21] Tsurtsumia, R. and Jordania, J., *Echoes from Georgia: Seventeen Arguments on Georgian Polyphony*, Nova Science Publishers, 2010.

[22] Erkvanidze, M., "The Georgian Musical System," in *6th International Workshop on Folk Music Analysis*, Dublin, Ireland, 2016.

[23] Tsereteli, Z. and Veshapidze, L., "On the Georgian traditional scale," pp. 288–295, Tbilisi, Georgia, 2014.

[24] Scherbaum, F., "On the benefit of Larynx-microphone field recordings for the documentation and analysis of polyphonic vocal music," in *6th International Workshop on Folk Music Analysis*, pp. 80–87, Dublin, Ireland, 2016.