Schubert Winterreise Dataset: A Multimodal Scenario for Music Analysis

CHRISTOF WEIß, FRANK ZALKOW, VLORA ARIFI-MÜLLER, and MEINARD MÜLLER, International Audio Laboratories Erlangen, Germany
HENDRIK VINCENT KOOPS and ANJA VOLK, Utrecht University, The Netherlands
HARALD G. GROHGANZ, Blue Square Group e. V., Germany

This article presents a multimodal dataset comprising various representations and annotations of Franz Schubert's song cycle *Winterreise*. Schubert's seminal work constitutes an outstanding example of the Romantic song cycle—a central genre within Western classical music. Our dataset unifies several public sources and annotations carefully created by music experts, compiled in a comprehensive and consistent way. The multimodal representations comprise the singer's lyrics, sheet music in different machine-readable formats, and audio recordings of nine performances, two of which are freely accessible for research purposes. By means of explicit musical measure positions, we establish a temporal alignment between the different representations, thus enabling a detailed comparison across different performances and modalities. Using these alignments, we provide for the different versions various musicological annotations describing tonal and structural characteristics. This metadata comprises chord annotations in different granularities, local and global annotations of musical keys, and segmentations into structural parts. From a technical perspective, the dataset allows for evaluating algorithmic approaches to tasks such as automated music transcription, cross-modal music alignment, or tonal analysis, and for testing these algorithms' robustness across songs, performances, and modalities. From a musicological perspective, the dataset enables the systematic study of Schubert's musical language and style in *Winterreise* and the comparison of annotations regarding different annotators and granularities. Beyond the research domain, the data may serve further purposes such as the didactic preparation of Schubert's work and its presentation to a wider public by means of an interactive multimedia experience. With this article, we provide a detailed description of the dataset, indicate its potential for computational music analysis by means of several studies, and point out possibilities for future research.

CCS Concepts: • **Information systems** → **Music retrieval**; *Content analysis and feature selection*; • **Applied computing** → **Sound and music computing**; *Digital libraries and archives*; *Document metadata*; • **Human-centered computing** → *Information visualization;*

Additional Key Words and Phrases: Dataset, music information retrieval, computational musicology, corpus analysis

ACM Reference format:
Christof Weiß, Frank Zalkow, Vlora Arifi-Müller, Meinard Müller, Hendrik Vincent Koops, Anja Volk, and Harald G. Grohganz. 2021. Schubert Winterreise Dataset: A Multimodal Scenario for Music Analysis. *J. Comput. Cult. Herit.* 14, 2, Article 25 (May 2021), 18 pages.
https://doi.org/10.1145/3429743

This work was supported by the German Research Foundation (DFG MU 2686/12-1).
Authors' addresses: C. Weiß, F. Zalkow, V. Arifi-Müller, and M. Müller, International Audio Laboratories Erlangen, Am Wolfsmantel 33, Erlangen, Germany, 91058; emails: {christof.weiss, frank.zalkow, vlora.arifi-mueller, meinard.mueller}@audiolabs-erlangen.de; H. V. Koops and A. Volk, Department of Information and Computing Sciences, Utrecht University, Princetonplein 5, De Uithof, Utrecht, The Netherlands, 3584 CC; emails: h.v.koops@gmail.com, a.volk@uu.nl; H. G. Grohganz, Blue Square Group e. V., Hermannstr. 18, 53225 Bonn; email: hg@blsq.org.

1556-4673/2021/05-ART25 $15.00
https://doi.org/10.1145/3429743

1 INTRODUCTION

With the advent of digital technology, musicology is being enriched with novel research methods. Digital archives and computational strategies enable musical corpus studies in a quantitative and systematic fashion [41, 44, 56, 61, 64, 69, 76]. Such studies have recently been performed based on a variety of different music representations including scanned sheet music images [61, 74], symbolic scores [5, 44, 73], MIDI files [35, 47, 79], and audio recordings [1, 41, 64, 75, 76].

For conducting corpus analyses, comprehensive, systematic, and reliable datasets are essential [70]. In general, there exist two strategies, each comprising two steps. In the first case, music experts analyze and annotate the primary material (e.g., scores or recordings) based on their music theory knowledge [11, 48] (*piece analysis*). Subsequently, computational and statistical methods are applied for systematically analyzing the annotations and drawing conclusions from these statistics [44] (*corpus analysis*). This approach entails full human control of the piece analysis but involves a decent amount of manual work, which becomes impractical in the case of large corpora. As an alternative approach, the primary material can be directly analyzed with (semi-)automatic methods, thus producing a computational or computer-assisted piece analysis, which serves as a basis for the corpus analysis [47, 61, 75, 76]. This second strategy requires reliable analysis algorithms, which need to be evaluated and—in the case of data-driven approaches—trained on high-quality, manually annotated datasets. In particular, machine-learning methods relying on deep neural networks require large amounts of data for training the networks [25].

This article presents a novel dataset with expert annotations, which can be useful in the context of both strategies and comprises the 24 songs of Franz Schubert's *Winterreise*, composed in 1827 for voice and piano. Concerning the primary material ("raw data"), the **Schubert Winterreise Dataset (SWD)** comprises several representations of the songs in different modalities (Figure 1 presents an overview). We provide the lyrics of the songs (text data) based on poems by Wilhem Müller ("poet" level). The dataset further contains sheet music ("composer" level) in the form of scans (image data) as well as machine-readable scores (symbolic data), manually created with the commercial software Sibelius (Avid Technologies, Burlington, MA). Along with the Sibelius source files, we provide exports to the MusicXML standard, which is readable by most music notation applications, and to the MIDI standard. For the "performer" level, we consider nine recorded performances (audio data). The two performances by singers Gerhard Hüsch (1933) and Randall Scarlata (2006) are directly provided within our dataset under a Creative Commons license. The further seven recordings, which we cannot publish due to copyright issues, are commercially available. We denote all of these manifestations of the songs as different *versions*, which are indicated by the horizontal axis in Figure 1.

Besides the raw data, our dataset contains secondary material (third dimension in Figure 1) comprising expert annotations of chords, local keys, global keys, and structural parts based on various sources [2, 26] and annotators. For example, we provide chord annotations using different levels of detail (chord vocabularies) [28], and local key analyses by three different annotators, which allows for studying harmonic ambiguity [34, 49].

To temporally link the different versions, we provide for the sheet music images and for the nine audio versions the positions of musical measures (downbeats), which are explicitly specified in the symbolic scores. Using the audio measure annotations as alignments refined with automatic synchronization techniques [20, 57, 58, 82], we transferred the analyses from the *musical* time axis of the scores (in measures) to the *physical* time axis of the recordings (in seconds) so that the SWD encompasses annotations directly referring to each of the nine audio versions, thus building up a three-dimensional "data tensor" as indicated by Figure 1.

As a multimodal, multi-version, and multi-annotator dataset, the SWD may be of high interest for the fields of music processing, music theory, and historical musicology. In particular, the data allows for studying diverse applications and tasks such as the mutual alignment between lyrics, score, MIDI, and audio data [3, 18, 23, 30, 31, 40, 46, 51, 59, 72], or between different audio versions [20, 57, 58, 82], the structural analysis of music [6, 27, 55], the tonal analysis of symbolic and audio data regarding global keys [54, 68], local keys [54, 83], chords

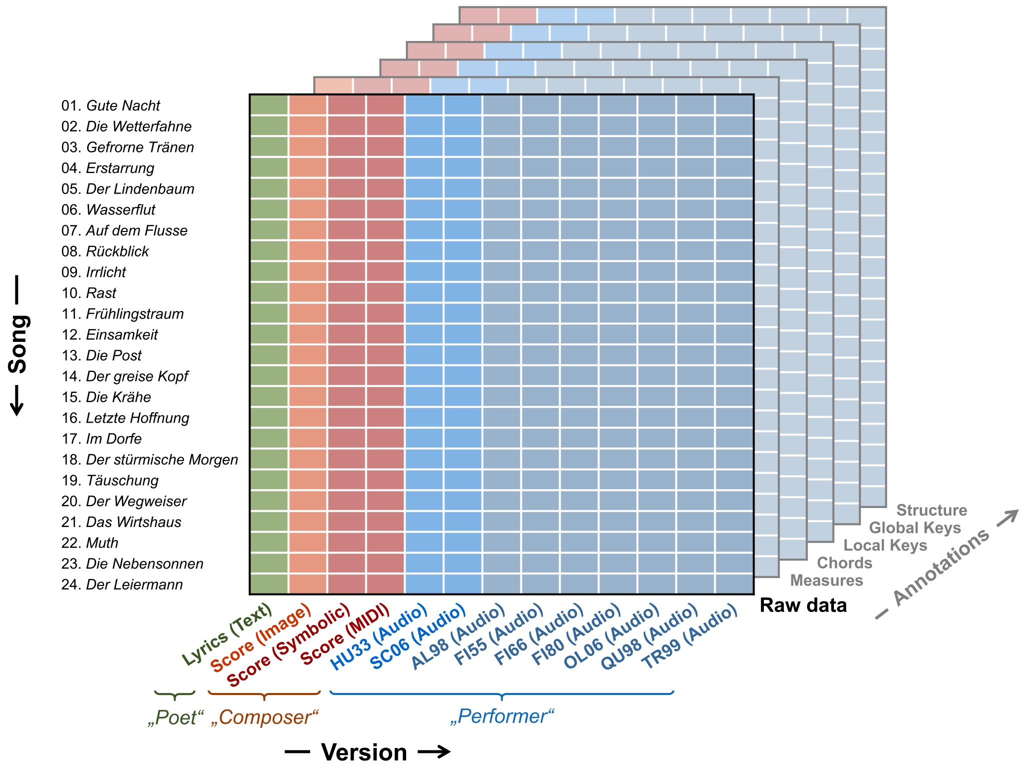


Fig. 1. Schematic overview of the SWD. For all 24 songs of *Winterreise*, the dataset comprises raw data in different representations (*versions*) such as lyrics, scores, or audio recordings of different performances. For the different versions, we provide time-aligned annotations of measure positions, chords, and local and global keys, as well as structural parts.

[12, 32, 42, 71], chord types [29, 76], and chord progressions [13, 15, 50, 52, 76], as well as the joint analysis of these aspects [39, 54, 60]. Some of these applications were already approached in previous studies [8, 26, 33, 45, 66, 77] using preliminary variants or subsets of the SWD. For the analysis tasks, the dataset allows for a systematic evaluation of the algorithms' robustness by comparing the results across different songs, versions, and annotations—sometimes referred to as *cross-version* experiments [16, 21, 32, 43, 78]. Figure 2 suggests possible splits of the dataset into subsets for training, validating, and testing machine-learning pipelines, as demonstrated by Schreiber et al. [66].

Beyond the study and development of technical methods, the data and annotations also provide an interesting source for systematic corpus analyses as mentioned earlier [44, 47, 61, 76]. In this context, the multimodality allows for a systematic comparison of the high-level corpus analysis results based on different input representations [75] and annotators [34, 49]. Further possible applications outside the research domain include didactical scenarios in schools or museums, where the various representations and annotations allow for creating interactive interfaces.

The remainder of this article is organized as follows. In Section 2, we give some background on *Winterreise* and an overview of the global structure and characteristics of the cycle. Section 3 describes the primary musical material published in the dataset. In Section 4, we explain the annotation process and the structure of the metadata. Section 5 highlights possible applications of the SWD by summarizing several previous publications

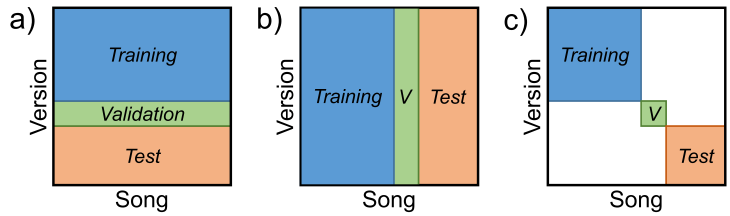


Fig. 2. Possibilities of splitting the SWD into training, validation, and test sets for machine-learning experiments. (a) Splitting along the version axis. (b) Splitting along the song axis. (c) Splitting along both axes. From Schreiber et al. [66].

and pointing out possible research directions. Finally, Section 6 concludes the article. The dataset is available as a version-controlled repository on the platform Zenodo with DOI: 10.5281/zenodo.4431535.[1]

2 SCHUBERT'S WINTERREISE

By the end of the 18th century, the *art song*—typically composed for solo voice with piano accompaniment based on secular lyrics and often denoted with the German term "Lied"—gained increasing popularity driven by composers such as Carl Friedrich Zelter or Johann Friedrich Reichardt ("Second Berlin School"). A few years later, it became the epitome of the upcoming Romantic period and its striving for a new naturalness in music. Pursuing this initiative, Franz Schubert (1797–1828) composed more than 600 such songs based on poetry and further developed the genre regarding, among other aspects, the musical form or the role of the piano accompaniment. In his later years, Schubert wrote several famous song cycles such as *Die schöne Müllerin* (1823), which are sequences of individual songs to be performed together, often forming a continuous narrative. In his most famous song cycle *Winterreise* (winter journey) D. 911/op. 89 (1827), a wandering protagonist tells his story of leaving town after his beloved falls for another. In the 24 songs (12 in each part), Schubert set poems by the German poet Wilhelm Müller (1794–1827) to music.

Table 1 shows an overview of the song cycle. Musically, the individual songs differ in length, structure, and complexity. Regarding their harmonic characteristics, some songs are rather unambiguous showing distinct key regions of diatonic pitch content (No. 2) or being based on a single tonic chord (No. 24). Other songs involve many altered chords (No. 10) and ambiguous key regions (No. 16). In general, minor keys prevail over major keys and slow tempi over fast ones. Schubert conceived the songs for "a singing voice with accompaniment of the piano." Since the lyrics indicate a male protagonist, male registers such as tenor and baritone are most popular. Although the original keys (Table 1) point to a high voice, score versions in various keys exist for the different registers (as is common for art songs), where male and female renderings differ by an octave shift in the singing part. In practice, singers even change between these versions from one song to another. As a consequence, the relationship between the songs' keys is not preserved.

The *Winterreise* cycle is one of the most-performed pieces within the classical music repertoire. For example, the classical-music streaming service Idagio[2] provides 137 performances (some of them incomplete) and the website GoOperalists 107 performances. Because of its high popularity and its outstanding artistic quality, *Winterreise* is of high interest for musicological and cultural research.

3 RAW DATA: MODALITIES AND VERSIONS

This section describes the creation and the format of the primary music sources ("raw data") within the SWD in the different modalities. Following Figure 1, we describe lyrics (text data), scores (image, symbolic, and MIDI

[1] https://doi.org/10.5281/zenodo.4431535.
[2] http://www.idagio.com (accessed November 21, 2019).

Table 1. Overview of Franz Schubert's Song Cycle *Winterreise* D. 911 Note: The 24 songs differ in key (referring to Schubert's original score), time signature, form, and length. Here, we only report the rough form with A' indicating a variant of A and parentheses indicating ambiguous forms

Song No.	Work ID	Title	Key (Score)	Time Sign.	Form	# Measures
1	D911-01	Gute Nacht	D minor	2/4	$A\,A\,A\,A'$	137
2	D911-02	Die Wetterfahne	A minor	6/8	$A\,B\,A'\,A'$	51
3	D911-03	Gefrorne Tränen	F minor	2/2	$A\,B\,C\,C$	55
4	D911-04	Erstarrung	C minor	4/4	$A\,B\,A'$	109
5	D911-05	Der Lindenbaum	E major	3/4	$A\,A'\,B\,A$	82
6	D911-06	Wasserflut	E minor	3/4	$A\,B\,A\,B$	60
7	D911-07	Auf dem Flusse	E minor	2/4	$A\,B\,C$	74
8	D911-08	Rückblick	G minor	3/4	$A\,B\,A'$	69
9	D911-09	Irrlicht	B minor	3/8	$A\,A\,B$	43
10	D911-10	Rast	C minor	2/4	$A\,B\,A\,B$	67
11	D911-11	Frühlingstraum	A major	6/8	$A\,B\,C\,A\,B\,C$	88
12	D911-12	Einsamkeit	B minor	2/4	$A\,B\,C\,C$	48
13	D911-13	Die Post	E♭ major	6/8	$A\,B\,A\,B$	94
14	D911-14	Der greise Kopf	C minor	3/4	$A\,B\,A'$	44
15	D911-15	Die Krähe	C minor	2/4	$A\,B\,A'$	43
16	D911-16	Letzte Hoffnung	E♭ major	3/4	$A\,B\,C\,D$	47
17	D911-17	Im Dorfe	D major	12/8	$A\,B\,A\,C$	49
18	D911-18	Der stürmische Morgen	D minor	4/4	$A\,B\,A'$	19
19	D911-19	Täuschung	A major	6/8	$A\,A\,B\,A$	43
20	D911-20	Der Wegweiser	G minor	2/4	$(A\,B\,A\,C)$	83
21	D911-21	Das Wirtshaus	F major	4/4	$(A\,A\,B\,C)$	31
22	D911-22	Mut	G minor	2/4	$A\,A\,B\,B$	64
23	D911-23	Die Nebensonnen	A major	3/4	$A\,A\,B\,A$	32
24	D911-24	Der Leiermann	A minor	3/4	$A\,A\,B$	61

data), and performances (audio data). Moreover, Figure 3 illustrates the different data types using song No. 14 from the SWD as an example.

3.1 Poet: Lyrics

As the first modality, we provide a textual representation (plain text files) of the lyrics in *Winterreise* (Figure 3(a)). The lyrics correspond to a series of poems by Wilhelm Müller (published 1823–24). The text files are structured into verses (by line breaks) and musical stanzas (by blank lines). Since the text is also part of the symbolic scores (Section 3.2), we do not provide separate musical time positions (measure/beat positions of words or syllables) within the isolated text files. Deviating from the original poems, we account for repeated and slightly modified verses or words so that the text files exactly correspond to the scores' lyrics.

3.2 Composer: Scores

On the "composer" level, we provide scores in different modalities. As the first modality, we include a graphical representation (image data) of the printed score in an edition by Peters [67], available as a scan at the International

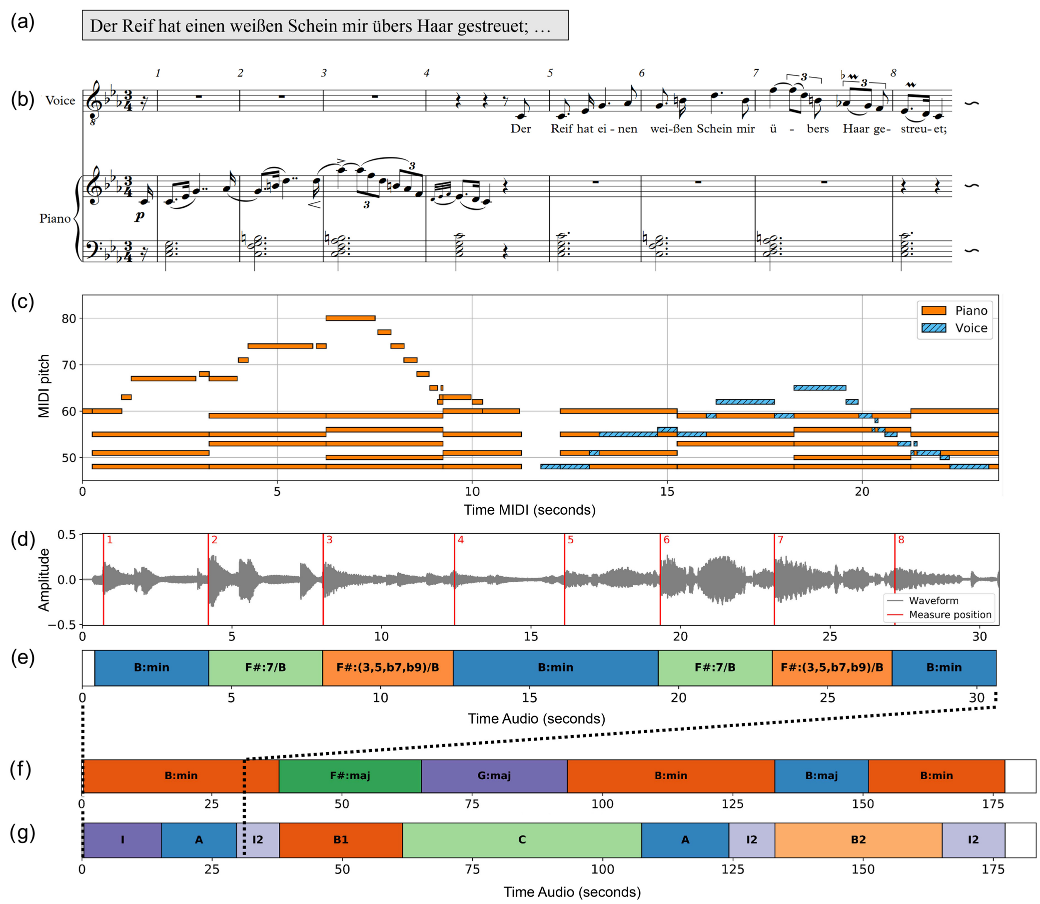


Fig. 3. Raw data in different modalities and several annotations for the first measures of song No. 14 "Der greise Kopf." (a) Lyrics (text data). (b) Score (symbolic data). (c) Piano roll (MIDI data). (d) Waveform (audio data) of the performance HU33 with annotated measure positions. (e) Chord annotations ("extended" vocabulary) aligned to the audio. (f) Local key annotations. (g) Structure annotations. Please note that (a) through (e) refer to an excerpt, whereas (f) and (g) refer to the full song, including silence at the end (white boxes).

Music Score Library Project (IMSLP).[3] We include a PDF file with additional markings of the measure numbers (68 pages). For the same scan, we provide individual PNG images of each page.

Next, we produced machine-readable symbolic scores (Figure 3(b)) by processing the scanned sheet music [67] with optical music recognition software (PhotoScore, Avid Technologies, Burlington, MA) and carefully correcting the results using commercial notation software (Sibelius, Avid Technologies). The symbolic scores follow the edition by Peters [67] mentioned earlier except for some minor notation variants (e.g., a different separation between left and right hand in the piano).[4] In song No. 1 "Gute Nacht," the first and second stanzas are notated using repetition signs. We unfold this repetition to simplify the assignment of lyrics to notes and to obtain a continuous measure count. For male singers, the treble clef is understood in a transposing way (down

[3] http://ks4.imslp.info/files/imglnks/usimg/9/92/IMSLP00414-Schubert_-_Winterreise.pdf.

[4] These variants do not play a role for the encoded music information (pitch, rhythm, etc.) but might affect symbol-level applications such as specific optical music recognition tasks. For instance, the appearance of dynamics or articulation symbols as well as measure or page numbers may vary.

Table 2. Recorded Performances in the SWD Note: For HU33 and SC06, we publish the audio data. All other recordings are commercially available and can be identified via the MusicBrainz release identifier. Durations are given in h:mm:ss

ID	Singer	Pianist	Year	Label	MusicBrainz ReleaseID	Dur.
HU33	Gerhard Hüsch	Hanns-Udo Müller	1933	*Public Domain*	9e2fcbe4-e7f3-45c2-b24e-eb304f261fa9	1:07:31
SC06	Randall Scarlata	Jeremy Denk	2006	*Creative Commons*	– *not on MusicBrainz* –	1:06:45
AL98	Thomas Allen	Roger Vignoles	1998	Virgin Classics, EMI	083c9ee1-77d7-4dea-9985-40f872607f59	1:13:33
FI55	Dietrich Fischer-Dieskau	Gerald Moore	1955	EMI Classics	06d71795-0a28-4c95-b802-d7525239c946	1:14:35
FI66	Dietrich Fischer-Dieskau	Jörg Demus	1966	Deutsche Gramm.	ae652682-2fa2-48c8-89f8-f3ff8a8cdccf	1:11:23
FI80	Dietrich Fischer-Dieskau	Daniel Barenboim	1980	Deutsche Gramm.	56a0e3e8-9c80-43ec-bd4e-4bcf599f9b2c	1:13:07
OL06	Thomas Oliemans	Bert van den Brink	2006	Fineline	– *not on MusicBrainz* –	1:14:42
QU98	Thomas Quasthoff	Charles Spencer	1998	RCA Red Seal (Sony)	f10034f7-5ebc-4122-a74d-35749ef90ca6	1:12:24
TR99	Roman Trekel	Ulrich Eisenlohr	1999	Naxos	26323984-c7d6-4ddb-ba46-5301254ddbca	1:15:21

one octave), whereas for female singers, it is realized as written. We provide the Sibelius source files together with exports to the standardized MusicXML format [24].

As a further score-like representation, we export MIDI files from Sibelius (Figure 3(c)), rendering the note events with a constant tempo. Voice and piano are separated into individual MIDI channels, with the voice corresponding to the male version (lower octave). Being aware of the availability of different historical score versions and editions, we confine ourselves to one manifestation of the score in the SWD. Nevertheless, our symbolic representations allows for an easy inclusion of notation variants and changes due to editorial decisions.

3.3 Performer: Audio Recordings

On the performer side, the SWD considers nine full performances of *Winterreise* (Table 2). Two of these performances are directly included in the dataset. The first one is a digitized vinyl record of a performance by baritone Gerhard Hüsch with pianist Hanns-Udo Müller from 1933 (denoted with the performance identifier HU33), which is now available under a Public Domain Mark 1.0 license[5] online.[6] Deviating from the score, the repetition in the first song (second stanza) is left out in this performance. To guarantee structural coherence with all other versions, we created a modified version of this song by duplicating the first stanza in the audio (the Python script for cutting is included in the dataset's extra material). This version now has the same number of measures and the same structure as all other versions—only the lyrics of the second stanza differ (identical to the first one). The original version of this song is included in the extra material of the SWD.

The second performance is a live recording performed by baritone Randall Scarlata with pianist Jeremy Denk (denoted as SC06), published under a Creative Commons Attribution Non-commercial No Derivatives 3.0 license[7] by the Isabella Stewart Gardner Museum (Boston).[8] Instead of the original part-wise audio files, we provide song-wise files as published on the METRUM project website.[9] Both performances are given as monaural audio files with a sampling rate of 22,050 Hz and a precision of 16 bits.

The SWD including all types of annotations extends to seven further performances. As is common for this song cycle, all singers are male, most of them baritones. The chosen keys vary slightly (in the range of a minor third) and the performances span roughly eight decades of recording history. Due to copyright issues, we cannot directly publish audio files for these performances. However, these recordings have been commercially available

[5] http://creativecommons.org/publicdomain/mark/1.0.
[6] http://musopen.org/de/music/8127-winterreise-d-911.
[7] http://creativecommons.org/licenses/by-nc-nd/3.0.
[8] https://www.gardnermuseum.org/experience/music?keys=winterreise+scarlata.
[9] http://www.audiolabs-erlangen.de/resources/MIR/METRUM-winterreise.

Fig. 4. Web application for synchronous playback with highlighting of measures in the score, using the sheet music images with annotations of measure bounding boxes.

at publication of this article. Table 2 provides an overview of all nine recordings together with release identifiers from MusicBrainz.[10]

4 ANNOTATIONS AND ANALYSES

In this section, we present the different types of annotations provided in the SWD and outline their creation process. The semantic annotations on the score level are manually created except for parts of the measure positions, which involved semi-automatic procedures. We then generated audio-based annotations using a semi-automatic transfer procedure.

4.1 Measure Positions

As outlined in Section 3, the SWD contains several versions of *Winterreise* in different modalities. For establishing temporal connections between these different versions, we provide musical measure annotations that indicate the corresponding spatial positions in the sheet music images (in pixels) and physical time positions in the audio recordings (in seconds).

For the sheet music images, our annotations specify the positions of the measures (all three staves) as rectangular bounding boxes indicating the coordinates of the upper left corner, the height, and the width. This data was generated with a set of tools for semi-automatic bounding box annotation [81] and is accessible online together with a web-based audio player,[11] which is shown in Figure 4.

Furthermore, we provide annotated measure or downbeat positions (Figure 3(d)) for all nine audio versions (Table 2). For the published versions (HU33 and SC06), these annotations were generated manually by carefully specifying each measure position using the Sonic Visualizer application [9]. We then transferred these measure annotations to the other seven performances by means of audio-to-audio synchronization techniques [20, 57, 58, 82]. In particular, we use the specific approach proposed by Zalkow et al. [82] to optimally choose between the versions HU33 and SC06 as the source version for each measure, utilizing the triple-based synchronization error [58] as an optimization criterion. Due to this elaborate process, the temporal precision of the measure positions is generally high with the annotations for HU33 and SC06 being most reliable (manually generated). Since the

[10]The CD information can be accessed via Music Brainz using the release identifiers from Table 2, e.g., https://musicbrainz.org/release/083c9ee1-77d7-4dea-9985-40f872607f59. Two of the recordings are not on MusicBrainz: OL06 can be found via the Amazon ASIN B000HLDD3S. SC06 is available online at the Isabella Stewart Gardner Museum (Boston) (https://www.gardnermuseum.org/experience/music?keys=winterreise+scarlata).
[11]http://www.audiolabs-erlangen.de/resources/MIR/2019-ISMIR-LBD-Measures.

measure positions were used to transfer the local annotations (see the following sections), these subsequently derived annotations are also more reliable for HU33 and SC06 than for the remaining versions.

4.2 Chords

In the following, we describe several types of tonal annotations provided in the SWD. First, we explain our annotations of musical chords (Figure 3(e)), which constitute a central notion within harmony analysis and are particularly relevant for describing the harmony in *Winterreise* since the piano most often plays a chordal accompaniment. The chord annotations are based on the detailed analyses published in the work of Absil [2], which comprise triad types, extensions such as sevenths, and additional information such as chord inversions together with a rough indication of the starting point for a sequence of chords. Since the exact temporal position of an individual chord in the score is not given by Absil [2], we complemented these annotations by specifying the exact boundaries of each chord segment in measures/beats and further performed some minor corrections to the chord labels.

Moreover, we translated the labels to the syntax specified by Harte et al. [28], which is common for chord annotations in **music information retrieval (MIR)** research. We provide these detailed annotations in the *shorthand* format as well as the *extended* format that specifies all chord tones as intervals over the root note [28]. Even though this notation is derived from pop/rock/jazz-style chord labels, it allows for a correct encoding of chords in Schubert's songs and can be parsed by common chord recognition and evaluation toolboxes. Whereas other harmony analysis datasets [11, 48] encode chords in their (key-dependent) harmonic function, we opt for a direct encoding of the *absolute* root note information. This way, we leave the chord annotations independent of the local keys, which allows for different annotation choices on the individual levels (see Section 4.3). Since many chord estimation systems in MIR use a restricted selection of chord types (denoted as "chord vocabulary"), we further provide simplifications of the chord labels to the 24 major and minor triads (*majmin*, based on the chord's third) and to major–minor with inversion information (*majmin_inv*).

We then transferred the chord annotations from the *musical* time axis of the scores (in measures and beats) to the *physical* time axis of the recordings (in seconds). For this purpose, we make use of the measure annotations (Section 4.1) and the global key annotations (Section 4.4). Additionally, we use a score-to-audio synchronization technique based on dynamic time warping and chroma features [20, 57] to obtain the precise beat positions within each measure, fixing the annotated measure positions as anchor points. With the resulting alignments, we transfer the segment boundaries of the score-based annotations to each audio version. We suitably transpose the labels using the global key annotations (Section 4.4) for score and audio versions, respectively. In Figure 5, we illustrate the annotation transfer in detail. From this procedure, we obtain for the score and each audio version song-wise CSV files specifying the segment boundaries (start and end) and the four different types of chord labels (*shorthand*, *extended*, *majmin*, and *majmin_inv*).

4.3 Local Keys

As another tonal information on a coarser temporal level, local specifications of the musical key (Figure 3(f)) are relevant for describing modulations and deriving harmonic functions of chords. In the SWD, we provide several such local key annotations based on different sources and annotators. Although all annotators are musical experts, the annotations vary since local key is a particularly ambiguous notion (referring to a *perceived* tonal center) and thus allow for studying subjectivity and ambiguity in harmonic annotations (as done for chord annotations in other works [34, 49]) and for evaluating the influence of annotations on automatic key detection experiments [77].

Since the dataset was created and improved in several cycles, these local key annotations are not independent of each other. Previous annotations were known to the annotators so that they carefully considered changing key labels based on their expertise and opinion. The first local key annotation (*ann1*) stems from the book [2],

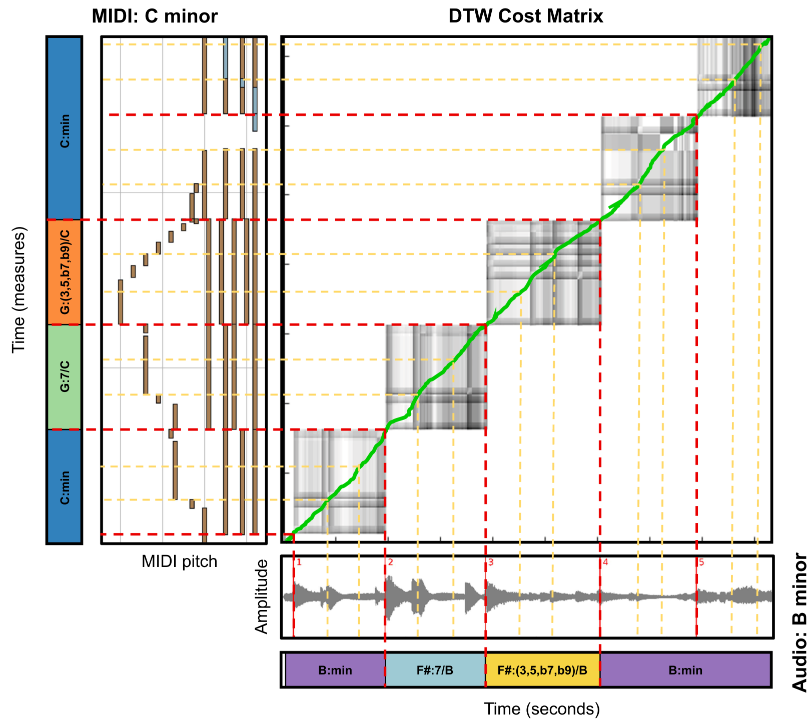


Fig. 5. Semi-automatic transfer of chord annotations for song No. 14 "Der greise Kopf" (see also Figure 3). The left-hand side represents the MIDI file with chord anntotations corresponding to the score (global key C minor). Using local MIDI-to-audio alignment with the measure annotations as anchor points (red dashed lines), we obtain detailed beat positions (dashed yellow lines) from the warping path (green line). With this alignment, we then transfer the chord annotations to the physical time axis of the audio recording and suitably transpose the chord labels to the performed key (B minor). DTW, dynamic time warping.

complemented with detailed measure boundaries of key segments. This annotation provides labels for all segments (including unisono and ambiguous passages) by always deciding on the most probable key. This follows the listener's attitude, who tends to always remember the last "clear" tonality [7, 37]. The second annotation (*ann2*) is based on Grohganz [26]. Being aware of *ann1*, this annotation avoids key labels for ambiguous segments by placing a "no key" label instead. The third annotation (*ann3*) by this article's first author relies on the previous two and also avoids "no key" regions [66]. In contrast to the others, this annotation aims for a clear distinction between local keys (as tonal centers) and chords (which might last for a considerable amount of time but still have a harmonic function toward a surrounding key[12]). Although we consider this annotation as the most reliable one, a comparison of the different annotations is highly interesting. All local key annotations are given as CSV files specifying start and end time (in measures or partial measures) as well as the local key label for each segment.

For transferring the local key annotations to the audio performances, we apply the same procedure as for the chord annotations (Section 4.2) using global key information for transposition and measure annotations as

[12]One such example is song song No. 12 "Einsamkeit," where a figurated C major chord (measures 31–32) can either be interpreted as a short modulation to the local key C major or as a Neapolitan chord within the following cadence in the local key B minor.

anchor points for score–audio alignment. The resulting local key annotations of the nine audio versions are given as CSV files indicating start and end time (in seconds) and local key label of each segment.

4.4 Global Keys

Regarding tonality, each song in *Winterreise* forms a self-contained entity with an established main key (global key). As mentioned in Section 2, singers prefer to individually choose a convenient key for each song using suitable transpositions. For this reason, we provide for all songs and performances information about the global key. These annotations are interesting for evaluating automatic key detection algorithms. Furthermore, the key information is crucial for transferring other tonal annotations such as chords (Section 4.2) or local keys (Section 4.3) to the audio versions. With the global key annotations for the performances and the global key of the score (Table 1), the necessary transposition of harmonic labels can be determined automatically (see Figure 5 for details). In the example of Figure 3, song No. 14 is notated in C minor in the score (Figure 3(b)) but the performance HU33 is in B minor so that the labels for chords (Figure 3(e)) and local keys (Figure 3(f)) have to be transposed by a suitable interval (here a semitone downward) when transferring score information to the audio.

4.5 Structure

Going beyond the description of tonality, we consider annotations of structural parts in the songs based on repetitions and variations. These annotations are based on a Ph.D. thesis [26] centered around music structure analysis, which used *Winterreise* as the main experimental scenario.[13] In this song cycle, we find different types of song structures such as the strophic song based on repetitions (*A A A*...), the variational strophic song (*A A1 A2*...), the ternary form (*A B A*), and the bar form (*A A B*). Furthermore, several songs are *through-composed* using different musical material for each of the lyrics' stanzas. In Table 1, we indicate the rough type of structure for each song with ambiguous segmentations in parentheses. Beyond the main structural parts (denoted with *A*, *B*, *C*, ...), we mark varied repetitions with additional numbers (*A1*, *A2*, ...). Instrumental parts such as an introduction or interlude are denoted by specific letters (*I*, *J*, *K*, ...).

For transferring the structure annotations to the audio versions, we again use the same procedure as for chords (Section 4.2) and local keys (Section 4.3). Figure 3(g) shows the structure annotation for a performance of song No. 14. For each song and version, we provide a CSV file with segment boundaries (start and end) and labels.

5 APPLICATIONS AND EXAMPLE ANALYSES

In the previous sections, we described the raw data and the annotations of the SWD, which stem from a number of sources, authors, and research projects. Many of these previous works already used parts or earlier versions of the *Winterreise* dataset for developing and evaluating (semi-)automatic approaches for synchronization, analysis, and visualization [8, 26, 33, 45, 66]. Summarizing some of these previous studies, this section illustrates the potential of the SWD for research in music processing and musicology as well as for further use cases.

Our first example is an early application of the *Winterreise* scenario for song retrieval and navigation in multimodal music databases based on lyrics search ("query-by-lyrics") [45]. At the core of this application is a cross-modal synchronization procedure, which breaks the lyrics–audio alignment down into two subproblems: lyrics–MIDI and MIDI–audio alignment. For the lyrics–MIDI alignment, the authors make use of a semi-automatic procedure that requires the isolated MIDI track of the singer and additional information about the melismatic/syllabic structure (how many notes correspond to each syllable). With the resulting alignment, the syllable positions in the MIDI file are determined. Then, the lyrics positions are automatically transferred to the audio using music synchronization techniques [46]. Figure 6 shows the SyncPlayer system [22]—an advanced audio player for multimodal presentation, browsing, and retrieval of music data. It allows the user to

[13] Visualizations and interactive applications for exploring the structure annotations can be found on the METRUM project website (http://www.audiolabs-erlangen.de/resources/MIR/METRUM-winterreise) and in the appendix of Grohganz [26].

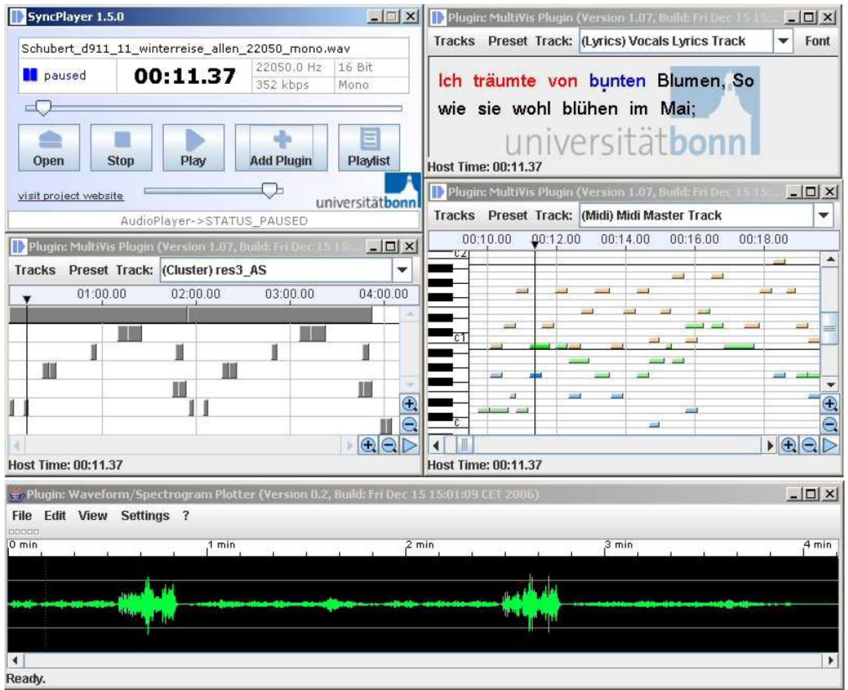
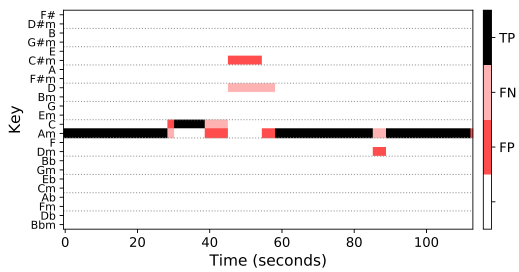


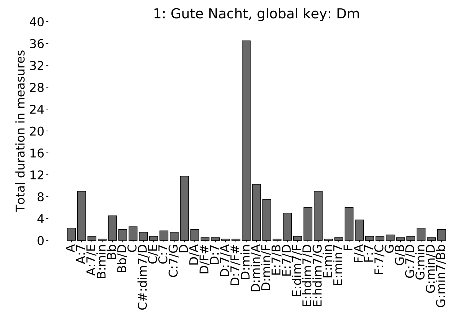
Fig. 6. Instance of the SyncPlayer interface for multimodal presentation, browsing, and retrieval of music data. The figure shows the main audio player (upper left) as well as plug-ins for visualizing lyrics (upper right), audio structure (middle left), piano-roll (middle right), and the waveform signal (bottom). From Müller et al. [45].

search for parts of the lyrics, retrieve performances of the matching songs, jump to the matching position within the performance, and synchronously play the performance from this position together with highlighted lyrics and piano rolls. Although the procedure applied here requires a suitable symbolic score with lyrics (aligned to score events) as provided by our MusicXML files, end-to-end systems for lyrics–audio alignment based on deep learning today show promising results [72]. These algorithms can be trained efficiently even with weakly aligned data on the word-, phrase-, or stanza level. For such approaches, the SWD contains an interesting training and evaluation scenario.

Next, we point to a recent study on local key estimation algorithms [66]. Based on the nine audio versions and the local key annotations (*ann3*) of the SWD, the authors compare a traditional system based on hidden Markov models and a modern system based on convolutional neural networks within different evaluation scenarios. Figure 7(a) shows the result of the convolutional neural network–based approach for song No. 15 (where keys are arranged according to the circle of fifths). As the main contribution, this article makes use of the different performances in the SWD and evaluates local key estimation systems in a cross-version setting. In particular, different ways of splitting the dataset into training, validation, and test data (Figure 2) are compared. Using several versions of the songs for training seems to suppress overfitting to irrelevant acoustic characteristics (known as album effect [53]). Further detailed results of this study can be found in the Ph.D. thesis [65]. With the different split possibilites, the SWD provides a great resource for systematically studying automatic analysis methods and their generalization capabilities, which becomes very important for building complex machine-learning systems. As an extension of this work, the output of the two local key estimation systems was recently



(a) Local key analysis of a machine-learning system for song No. 15 “Die Krähe” (QU98). Dark red bars indicate false positive predictions, bright red bars false negatives, black bars true positives. Figure from [66].



(b) Statistics of different chords in song No. 1 “Gute Nacht” based on the total duration of the chords as annotated in the SWD. Figure from [33].

Fig. 7. Application of the SWD for analyzing local keys (a) and chords (b).

analyzed in the light of annotator subjectivity, making use of the multiple annotations within the SWD introduced in Section 4.3.

Similar applications can also be realized for automatic chord estimation. In the work of Koops [33, Chapter 5], such a cross-version experiment was performed on the nine audio performances using the chord annotations of the SWD in the *majmin* vocabulary. The thesis chapter provides a detailed introduction of the *Winterreise* cycle and the chord annotations by showing interesting statistics about chords and chord types (Figure 7(b)). In the experiments, different systems are compared, showing that a transfer from deep neural networks trained on popular music to the SWD scenario is challenging and that results can vary greatly across versions. Statistics shown in the work of Koop [33] indicate that major and minor triads account for only 66% of audio duration, whereas dominant seventh chords and different types of diminished chords cover further 25%. For this reason, further research on automatic chord estimation in the SWD scenario (and classical music in general) using larger chord vocabularies [14, 42] is of high interest for MIR and machine-learning research.

Finally, we want to illustrate the application of the SWD for music structure analysis. In the Ph.D. thesis [26], the *Winterreise* cycle served as a central scenario for segmentation tasks according to the principles of repetition (melody) and homogeneity (local key regions). In particular, techniques from the field of image processing were applied to enhance path and block structures in self-similarity matrices. Besides the automatic segmentation, hierachical visualizations based on *scape plots* [62, 63] turned out to be beneficial for presenting harmonic and structural analyses. Figure 8 shows the interactive display of such scape plots on the METRUM project website[14] including a double-scape plot (rhombus plot) that indicates major-key (upper triangle) and minor-key (lower triangle) results for each position and segment length. The color scheme is inspired by the circle of fifths with neighboring keys sharing similar colors. Black indicates no shift, red a shift toward upper-fifth or sharp keys, and blue toward lower-fifth or flat keys. Green colors are used for more distant keys, and cyan and yellow correspond to the relative key as well.

These visual interfaces were reconsidered in a recent student thesis [8], which provides novel visualization techniques and compares the results for the audio version HU33 and SC06 considering their different global keys. The structure analyses and visualizations show that the SWD can be useful not only for improving automatic analysis and segmentation systems but also for developing novel types of interfaces, which allow musicologists to gain new insights into the characteristics of musical pieces—even for situations of ambiguous harmony or structure. Beyond the research domain, such visualizations and interfaces bear great potential for other fields. For example, the didactic preparation of the song cycle for museums or schools can benefit from multimedial and

[14] http://www.audiolabs-erlangen.de/resources/MIR/METRUM-winterreise.

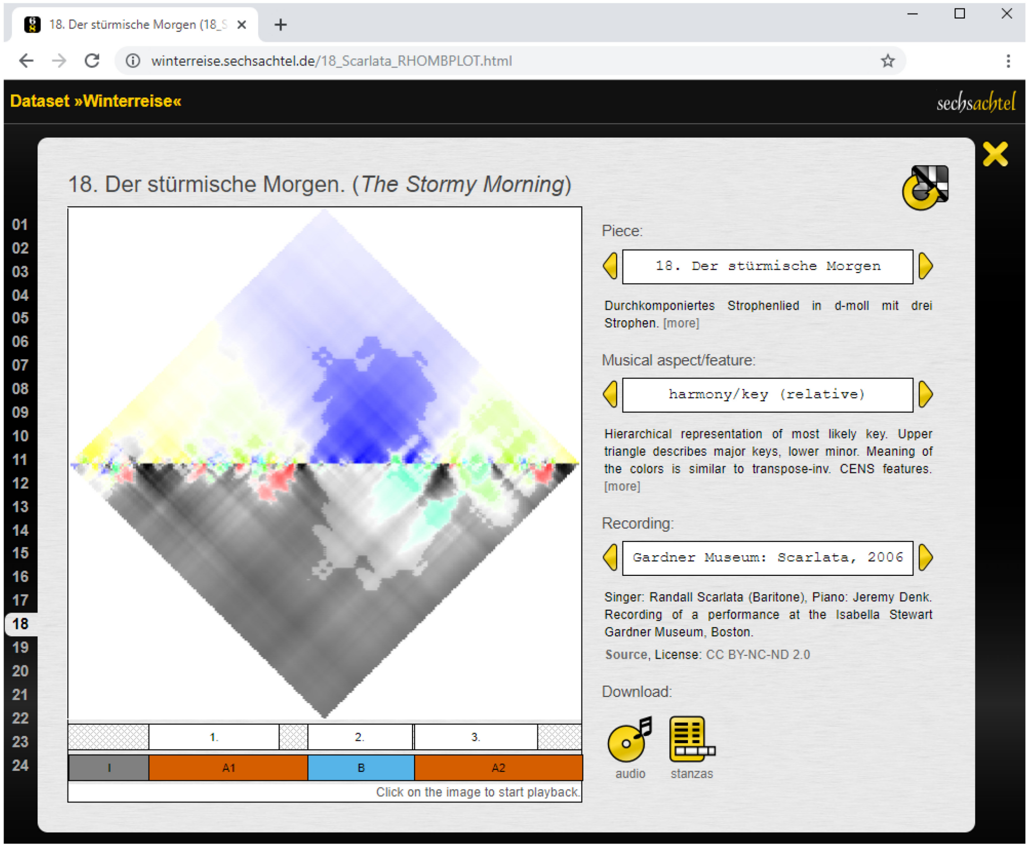


Fig. 8. Automated local key analysis for song No. 18 "Der stürmische Morgen," visualized in a double-scape or rhombus plot with major keys in the upper triangle and minor keys in the lower one [26]. This is a screenshot of a web application that allows for visualization of analyses and annotations with synchronous playback of the underlying audio version.

interactive experiences. Regarding commercial application, dedicated classical music streaming services such as Idagio[15] and Primephonic[16] may exploit datasets such as the SWD for enhancing their users' listening experience. Moreover, even live performances may be enriched with additional information presented in a multimedial way, as done for example in the PHENICX research project [38].

Altogether, these previous experiments show that *Winterreise* constitutes a complex music scenario, which is of high interest for studying different music processing tasks and applications. However, these studies have only taken first steps toward developing fully fledged analysis systems and interfaces that can be used in real-world applications. Besides the tasks presented in this section, many further applications can be studied on the basis of the SWD. One example is the end-to-end cross-modal alignment of text, image, and audio data, which has recently been approached for the tasks of lyrics–audio alignment [72], score following in symbolic scores [4] and in sheet music images [17], and for finding image–audio correspondences [18]. A further application possibility is the joint analysis of (down)beats, keys, chords, and structure, which has been considered in several publications [39, 54, 60]. Finally, the SWD may enable interesting research possibilities for the whole field of

[15] http://www.idagio.com.
[16] http://www.primephonic.com/.

music performance analysis [10, 36, 80]. In particular, the variety of audio recordings in the dataset allows for systematically studying performance characteristics such as loudness [36], expressive timing [10], or vibrato [19].

6 CONCLUSION

In this article, we presented a multimodal dataset comprising different music representations and annotations of Franz Schubert's song cycle *Winterreise*. The primary material (raw data) consists of textual representations of the songs' lyrics, music scores in image, symbolic, and MIDI format, as well as nine audio recordings of performances. The secondary material (annotations) comprises information about musical measure positions in sheet music images and audio recordings, as well as expert analyses of chords, local keys, global keys, and structural parts. Relying on the measure positions, we semi-automatically transferred the manual annotations across the different modalities and versions. The resulting annotated dataset allows for studying music processing tasks in a cross-version setting. In particular, algorithms for audio segmentation, structural analysis, lyrics alignment, or tonal analysis can be investigated regarding their capabilities for generalizing across songs, versions, and modalities. By referring to several previous publications, we discussed some of these manifold research opportunities provided by the SWD. Regarding future work, the SWD may facilitate the musicological analysis of the *Winterreise* cycle, the development of (semi-)automatic synchronization, analysis, and retrieval systems, and the realization of interactive software and visualization applications, which bear potential also for didactical and commercial applications. Thus, the dataset may trigger and enrich interesting research directions within the fields of music processing, computational musicology, and cultural heritage.

ACKNOWLEDGMENTS

The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS. We thank Junseong Park, Moritz Berendes, Julia Strahwald, Selina Sonntag, and all other students involved in the preparation of data and annotations. Furthermore, we thank Dr. Frans Absil for providing textual sources for the chord annotations.

REFERENCES

[1] Jakob Abeßer, Klaus Frieler, Estefanía Cano, Martin Pfleiderer, and Wolf-Georg Zaddach. 2017. Score-informed analysis of tuning, intonation, pitch modulation, and dynamics in jazz solos. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25, 1 (2017), 168–177.

[2] Frans G. J. Absil. 2017. *Musical Analysis—Visiting the Great Composers* (6th ed.). Frans Absil Music.

[3] Andreas Arzt. 2016. *Flexible and Robust Music Tracking.* Ph.D. Dissertation. Universität Linz, Austria.

[4] Andreas Arzt and Gerhard Widmer. 2015. Real-time music tracking using multiple performances as a reference. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 357–363.

[5] Héctor G. Bellmann. 2012. *Categorization of Tonal Music Style: A Quantitative Investigation.* Ph.D. Dissertation. Griffith University, Brisbane, Australia.

[6] Juan Pablo Bello. 2011. Measuring structural similarity in music. *IEEE Transactions on Audio, Speech, and Language Processing* 19, 7 (2011), 2013–2025.

[7] Anthony J. Bishara and Gabriel A. Radvansky. 2005. The detection and tracing of melodic key changes. *Perception & Psychophysics* 67, 1 (2005), 36–47.

[8] Leo Brütting. 2019. *Hierarchical Tonal Analysis of Music Signals.* Bachelor's Thesis. Friedrich-Alexander-University of Erlangen-Nuremberg (FAU), Germany.

[9] 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.* 1467–1468.

[10] Carlos Eduardo Cancino Chacón, Maarten Grachten, Werner Goebl, and Gerhard Widmer. 2018. Computational models of expressive music performance: A comprehensive and critical review. *Frontiers in Digital Humanities* 5 (2018), 25.

[11] Tsung-Ping Chen and Li Su. 2018. Functional harmony recognition of symbolic music data with multi-task recurrent neural networks. In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR).* 90–97.

[12] Taemin Cho and Juan Pablo Bello. 2014. On the relative importance of individual components of chord recognition systems. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 22, 2 (2014), 477–492.

[13] W. Bas de Haas, Remco C. Veltkamp, and Frans Wiering. 2008. Tonal pitch step distance: A similarity measure for chord progressions. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 51–56.
[14] Jun-Qi Deng and Yu-Kwong Kwok. 2016. A hybrid Gaussian-HMM-deep learning approach for automatic chord estimation with very large vocabulary. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 812–818.
[15] Bruno Di Giorgi, Simon Dixon, Massimiliano Zanoni, and Augusto Sarti. 2017. A data-driven model of tonal chord sequence complexity. *IEEE/ACM Transactions on Audio, Speech, and Language Processing* 25, 11 (2017), 2237–2250.
[16] Christian Dittmar, Bernhard Lehner, Thomas Prätzlich, Meinard Müller, and Gerhard Widmer. 2015. Cross-version singing voice detection in classical opera recordings. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 618–624.
[17] Matthias Dorfer, Andreas Arzt, and Gerhard Widmer. 2016. Towards score following in sheet music images. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 789–795.
[18] Matthias Dorfer, Andreas Arzt, and Gerhard Widmer. 2017. Learning audio-sheet music correspondences for score identification and offline alignment. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 115–122.
[19] Jonathan Driedger, Stefan Balke, Sebastian Ewert, and Meinard Müller. 2016. Template-based vibrato analysis of music signals. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 239–245.
[20] Sebastian Ewert, Meinard Müller, and Peter Grosche. 2009. High resolution audio synchronization using chroma onset features. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).* 1869–1872.
[21] Sebastian Ewert, Meinard Müller, Verena Konz, Daniel Müllensiefen, and Gerraint A. Wiggins. 2012. Towards cross-version harmonic analysis of music. *IEEE Transactions on Multimedia* 14, 3-2 (2012), 770–782.
[22] Christian Fremerey, Frank Kurth, Meinard Müller, and Michael Clausen. 2007. A demonstration of the SyncPlayer system. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 131–132.
[23] Hiromasa Fujihara, Masataka Goto, Jun Ogata, and Hiroshi G. Okuno. 2011. LyricSynchronizer: Automatic synchronization system between musical audio signals and lyrics. *IEEE Journal of Selected Topics in Signal Processing* 5, 6 (2011), 1252–1261.
[24] Michael Good. 2001. MusicXML for notation and analysis. *Computing in Musicology* 12 (2001), 113–124.
[25] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning.* MIT Press, Cambridge, MA.
[26] Harald Grohganz. 2015. *Algorithmen zur strukturellen analyse von musikaufnahmen.* Ph.D. Dissertation. University of Bonn, Germany.
[27] Harald Grohganz, Michael Clausen, Nanzhu Jiang, and Meinard Müller. 2013. Converting path structures into block structures using eigenvalue decompositions of self-similarity matrices. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 209–214.
[28] Christopher Harte, Mark B. Sandler, Samer Abdallah, and Emilia Gómez. 2005. Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 66–71.
[29] Aline Honingh and Rens Bod. 2010. Pitch class set categories as analysis tools for degrees of tonality. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 459–464.
[30] Ning Hu, Roger B. Dannenberg, and George Tzanetakis. 2003. Polyphonic audio matching and alignment for music retrieval. In *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).*
[31] Min-Yen Kan, Ye Wang, Denny Iskandar, Tin Lay Nwe, and Arun Shenoy. 2008. LyricAlly: Automatic synchronization of textual lyrics to acoustic music signals. *IEEE Transactions on Audio, Speech, and Language Processing* 16, 2 (2008), 338–349.
[32] Verena Konz, Meinard Müller, and Rainer Kleinertz. 2013. A cross-version chord labelling approach for exploring harmonic structures—A case study on Beethoven's appassionata. *Journal of New Music Research* 42, 1 (2013), 61–77.
[33] Hendrik Vincent Koops. 2019. *Computational Modelling of Variance in Musical Harmony.* Ph.D. Dissertation. Utrecht University, The Netherlands.
[34] Hendrik Vincent Koops, W. Bas de Haas, John Ashley Burgoyne, Jeroen Bransen, Anna Kent-Muller, and Anja Volk. 2019. Annotator subjectivity in harmony annotations of popular music. *Journal of New Music Research* 48, 3 (2019), 232–252.
[35] Hendrik Vincent Koops, Anja Volk, and W. Bas de Haas. 2015. Corpus-based rhythmic pattern analysis of ragtime syncopation. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 483–489.
[36] Katerina Kosta, Oscar F. Bandtlow, and Elaine Chew. 2014. Practical implications of dynamic markings in the score: Is piano always piano? In *Proceedings of the AES International Conference on Semantic Audio.*
[37] Carol L. Krumhansl and Edward J. Kessler. 1982. Tracing the dynamic changes in perceived tonal organization in a spatial representation of musical keys. *Psychological Review* 89, 4 (1982), 334–368.
[38] Cynthia C. S. Liem, Emilia Gómez, and Markus Schedl. 2015. PHENICX: Innovating the classical music experience. In *Proceedings of the IEEE International Conference on Multimedia and Expo Workshops (ICMEW).*
[39] Matthias Mauch and Simon Dixon. 2010. Simultaneous estimation of chords and musical context from audio. *IEEE Transactions on Audio, Speech, and Language Processing* 18, 6 (2010), 1280–1289.
[40] Matthias Mauch, Hiromasa Fujihara, and Masataka Goto. 2012. Integrating additional chord information into HMM-based lyrics-to-audio alignment. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 1 (2012), 200–210.
[41] Matthias Mauch, Robert M. MacCallum, Mark Levy, and Armand M. Leroi. 2015. The evolution of popular music: USA 1960–2010. *Royal Society Open Science* 2, 5 (2015), 150081.

[42] Brian McFee and Juan Pablo Bello. 2017. Structured training for large-vocabulary chord recognition. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 188–194.
[43] Stylianos I. Mimilakis, Christof Weiss, Vlora Arifi-Müller, Jakob Abesser, and Meinard Müller. 2019. Cross-version singing voice detection in opera recordings: Challenges for supervised learning. In *Machine Learning and Knowledge Discovery in Databases*. Communications in Computer and Information Science, Vol. 1168. Springer, 429–436.
[44] Fabian C. Moss, Markus Neuwirth, Daniel Harasim, and Martin Rohrmeier. 2019. Statistical characteristics of tonal harmony: A corpus study of Beethovenś string quartets. *PLOS ONE* 14, 6 (2019), e0217242.
[45] Meinard Müller, Frank Kurth, David Damm, Christian Fremerey, and Michael Clausen. 2007. Lyrics-based audio retrieval and multimodal navigation in music collections. In *Proceedings of the European Conference on Digital Libraries (ECDL)*. 112–123.
[46] Meinard Müller, Frank Kurth, and Tido Röder. 2004. Towards an efficient algorithm for automatic score-to-audio synchronization. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 365–372.
[47] Eita Nakamura and Kunihiko Kaneko. 2019. Statistical evolutionary laws in music styles. *Scientific Reports* 9, 1 (2019), 15993.
[48] Markus Neuwirth, Daniel Harasim, Fabian C. Moss, and Martin Rohrmeier. 2018. The annotated Beethoven corpus (ABC): A dataset of harmonic analyses of all Beethoven string quartets. *Frontiers in Digital Humanities* 5 (2018), 16.
[49] Yizhao Ni, Matt McVicar, Raúl Santos-Rodríguez, and Tijl De Bie. 2013. Understanding effects of subjectivity in measuring chord estimation accuracy. *IEEE Transactions on Audio, Speech, and Language Processing* 21, 12 (2013), 2607–2615.
[50] Mitsunori Ogihara and Tao Li. 2008. N-gram chord profiles for composer style identification. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 671–676.
[51] Nicola Orio and Diemo Schwarz. 2001. Alignment of monophonic and polyphonic music to a score. In *Proceedings of the International Computer Music Conference (ICMC)*. 155–158.
[52] Jean-François Paiement, Douglas Eck, and Samy Bengio. 2005. A probabilistic model for chord progressions. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 312–319.
[53] Elias Pampalk, Arthur Flexer, and Gerhard Widmer. 2005. Improvements of audio-based music similarity and genre classification. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 628–633.
[54] Hélène Papadopoulos and Geoffroy Peeters. 2012. Local key estimation from an audio signal relying on harmonic and metrical structures. *IEEE Transactions on Audio, Speech, and Language Processing* 20, 4 (2012), 1297–1312.
[55] Jouni Paulus, Meinard Müller, and Anssi Klapuri. 2010. Audio-based music structure analysis. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 625–636.
[56] Martin Pfleiderer, Klaus Frieler, Jakob Abesser, Wolf-Georg Zaddach, and Benjamin Burkhart. 2017. *Inside the Jazzomat*. Schott Campus, Mainz, Germany.
[57] Thomas Prätzlich, Jonathan Driedger, and Meinard Müller. 2016. Memory-restricted multiscale dynamic time warping. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 569–573.
[58] Thomas Prätzlich and Meinard Müller. 2016. Triple-based analysis of music alignments without the need of ground-truth annotations. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 266–270.
[59] Christopher Raphael. 2006. Aligning music audio with symbolic scores using a hybrid graphical model. *Machine Learning* 65, 2-3 (2006), 389–409.
[60] Thomas Rocher, Matthias Robine, Pierre Hanna, and Laurent Oudre. 2010. Concurrent estimation of chords and keys from audio. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*. 141–146.
[61] Pablo H. Rodriguez Zivic, Favio Shifres, and Guillermo A. Cecchi. 2013. Perceptual basis of evolving Western musical styles. *Proceedings of the National Academy of Sciences* 110, 24 (2013), 10034–10038.
[62] Craig Stuart Sapp. 2001. Harmonic visualizations of tonal music. In *Proceedings of the International Computer Music Conference (ICMC)*. 423–430.
[63] Craig Stuart Sapp. 2005. Visual hierarchical key analysis. *ACM Computers in Entertainment* 3, 4 (2005), 1–19.
[64] 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*. 29–36.
[65] Hendrik Schreiber. 2020. *Data-Driven Approaches for Tempo and Key Estimation of Music Recordings*. Ph.D. Dissertation. Friedrich-Alexander-University of Erlangen-Nuremberg (FAU), Germany.
[66] Hendrik Schreiber, Christof Weiss, and Meinard Müller. 2020. Local key estimation in classical music recordings: A cross-version study on Schubert's Winterreise. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*. 501–505.
[67] Franz Schubert. 1827. Winterreise. Ein Cyclus von Liedern von Wilhelm Müller. Retrieved March 17, 2021 from http://ks4.imslp.info/files/imglnks/usimg/9/92/IMSLP00414-Schubert_-_Winterreise.pdf. Gesänge für eine Singstimme mit Klavierbegleitung, Edition Peters, No.20a, n.d. Plate 9023.
[68] Björn Schuller and Benedikt Gollan. 2012. Music theoretic and perception-based features for audio key determination. *Journal of New Music Research* 41, 2 (2012), 175–193.

[69] Xavier Serra. 2014. Computational approaches to the art music traditions of India and Turkey. *Journal of New Music Research: Special Issue on Computational Approaches to the Art Music Traditions of India and Turkey* 43, 1 (2014), 1–2.
[70] 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*.
[71] Siddharth Sigtia, Nicolas Boulanger-Lewandowski, and Simon Dixon. 2015. Audio chord recognition with a hybrid recurrent neural network. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 127–133.
[72] Daniel Stoller, Simon Durand, and Sebastian Ewert. 2019. End-to-end lyrics alignment for polyphonic music using an audio-to-character recognition model. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP).* 181–185.
[73] David Temperley. 1997. An algorithm for harmonic analysis. *Music Perception* 15, 1 (1997), 31–68.
[74] Vladimir Viro. 2011. Peachnote: Music score search and analysis platform. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR).* 359–362.
[75] Christof Weiss, Stefan Balke, Jakob Abesser, and Meinard Müller. 2018. Computational corpus analysis: A case study on jazz solos. In *Proceedings of the 19th International Society for Music Information Retrieval Conference (ISMIR).* 416–423.
[76] Christof Weiss, Matthias Mauch, Simon Dixon, and Meinard Müller. 2019. Investigating style evolution of Western classical music: A computational approach. *Musicae Scientiae* 23, 4 (2019), 486–507.
[77] Christof Weiss, Hendrik Schreiber, and Meinard Müller. 2020. Local key estimation in music recordings: A case study across songs, versions, and annotators. *IEEE/ACM Transactions on Audio, Speech & Language Processing* 28 (2020), 2919–2932.
[78] Christof Weiss, Frank Zalkow, Meinard Müller, Stephanie Klauk, and Rainer Kleinertz. 2017. Versionsübergreifende visualisierung harmonischer verläufe: Eine fallstudie zu wagners ring-zyklus. In *Proceedings of the Jahrestagung der Gesellschaft für Informatik (GI).* 205–217.
[79] Christopher Wm. White. 2013. *Some Statistical Properties of Tonality, 1650–1900.* Ph.D. Dissertation. Yale University, New Haven, CT.
[80] Gerhard Widmer, Simon Dixon, Werner Goebl, Elias Pampalk, and Asmir Tobudic. 2003. In search of the Horowitz factor. *AI Magazine* 24, 3 (2003), 111–130.
[81] Frank Zalkow, Angel Villar Corrales, T. J. Tsai, Vlora Arifi-Müller, and Meinard Müller. 2019. Tools for semi-automatic bounding box annotation of musical measures in sheet music. In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR).*
[82] Frank Zalkow, Christof Weiss, Thomas Prätzlich, Vlora Arifi-Müller, and Meinard Müller. 2017. A multi-version approach for transferring measure annotations between music recordings. In *Proceedings of the AES International Conference on Semantic Audio.* 148–155.
[83] Yongwei Zhu and Mohan S. Kankanhalli. 2004. Key-based melody segmentation for popular songs. In *Proceedings of the International Conference on Pattern Recognition (ICPR)*, Vol. 3. 862–865.

Received December 2019; revised July 2020; accepted October 2020