

DATASET

Dagstuhl ChoirSet: A Multitrack Dataset for MIR Research on Choral Singing

Sebastian Rosenzweig*, Helena Cuesta†, Christof Weiß*, Frank Scherbaum‡, Emilia Gómez†,§ and Meinard Müller*

Choral singing is a central part of musical cultures across the world, yet many facets of this widespread form of polyphonic singing are still to be explored. Music information retrieval (MIR) research on choral singing benefits from multitrack recordings of the individual singing voices. However, there exist only few publicly available multitrack datasets on polyphonic singing. In this paper, we present Dagstuhl ChoirSet (DCS), a multitrack dataset of a cappella choral music designed to support MIR research on choral singing. The dataset includes recordings of an amateur vocal ensemble performing two choir pieces in full choir and quartet settings. The audio data was recorded during an MIR seminar at Schloss Dagstuhl using different close-up microphones to capture the individual singers' voices. In this article, we give detailed insights into all stages of creating DCS: recording process, data preparation, generation of annotations as well as development of suitable interfaces for publicly accessing and reusing the data. Furthermore, we demonstrate the potential of the dataset for MIR research by discussing case studies on choral intonation assessment and multiple fundamental frequency (F0) estimation.

Keywords: Dagstuhl ChoirSet; Choral Music; Multitrack Recording; Choral Singing Analysis; FO

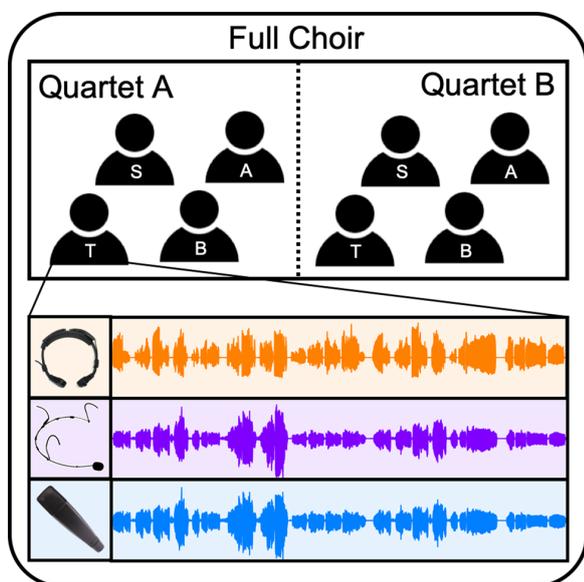


Figure 1: Dagstuhl ChoirSet—an overview.

* International Audio Laboratories Erlangen, DE

† Music Technology Group, Universitat Pompeu Fabra, Barcelona, ES

‡ University of Potsdam, DE

§ Joint Research Centre, European Commission, Seville, ES

Corresponding author: Sebastian Rosenzweig
(sebastian.rosenzweig@audiolabs-erlangen.de)

1. Introduction

Choral singing is one of the most widespread types of polyphonic singing (Sundberg, 1987). For instance, the European Choral Association¹ reports over 37 million amateur and professional choir singers on the European continent, while Chorus America² reports 54 million active singers in the U.S. The great interest in choral singing motivates the need for MIR technologies to support singers and conductors in their rehearsal practices (Gómez et al., 2020) via mobile applications^{3,4} and web-based interfaces.⁵ Over the last years, there has been an increasing number of MIR techniques developed for analyzing polyphonic vocal music (Dai and Dixon, 2017; Mauch et al., 2014; Cuesta et al., 2018; Devaney, 2011; Devaney and Ellis, 2008; Howard et al., 2013; Howard, 2007; Weiß et al., 2019) as well as for synthesizing expressive singing (Chandna et al., 2019; Blaauw and Bonada, 2017). Essential to the development of such techniques is the availability of suitable datasets and processing tools. In particular, multitrack recordings are of great value for evaluation purposes. However, due to high demands on recording equipment and infrastructure, there exist only few publicly available multitrack datasets on polyphonic vocal music.

The lack of suitable research data was one of the driving motivations to create *Dagstuhl ChoirSet* (DCS), a publicly available multitrack dataset of a cappella choral music for MIR research (cf. **Figure 1**). The audio data was recorded

during a one-week research seminar on “Computational Methods for Melody and Voice Processing in Music Recordings” (Müller et al., 2019) at Schloss Dagstuhl.⁶ For the recordings, we assembled a vocal ensemble of mostly amateur singers (all were participants of the Dagstuhl seminar) covering different SATB (Soprano, Alto, Tenor, and Bass) voice sections. After several rehearsals with a conductor, we recorded multiple takes of two choir pieces in a full choir setting and two quartet settings (Quartet A and Quartet B). Furthermore, we recorded some systematic exercises for practicing choral intonation. As one main feature of the dataset, individual singers were recorded using multiple close-up microphones, including larynx, headset, and dynamic microphones. Subsequent to recording and curating the recorded multitrack data, we annotated beat positions and generated time-aligned score representations for each of the music recordings. Furthermore, we automatically extracted F0-trajectories for all close-up microphone signals. The publicly available dataset is archived on Zenodo⁷ and is accessible via an interactive web-based interface with score-following and playback functionality.⁸ In order to facilitate reproducibility and further research using this dataset, we have created an open source Python toolbox with helper functions to load, parse, and process dataset files.⁹

In summary, our annotated dataset has different musical and acoustical dimensions that open up a variety of research scenarios. Besides being a good basis for studying amateur choral singing, DCS constitutes a challenging scenario for various fundamental tasks in MIR such as automatic music transcription (Benetos et al., 2019), score-to-audio alignment (Thomas et al., 2012), and beat tracking (Zapata et al., 2014; Böck et al., 2019). Moreover, the close-up microphone signals as well as the available F0-trajectories and scores can serve as a baseline to research on (informed) source separation techniques (Cano et al., 2019, 2014). Furthermore, it allows for comparisons between multiple choir/quartet performances, choir settings, and microphone types.

The remainder of this article is structured as follows. In Section 2, we give an overview on datasets related to our

work. In Section 3, we describe DCS by providing details on the choir settings, selected pieces, technical setup of the recordings, and generated annotations. In Section 4, we explain the different interfaces to access and use the dataset. In Section 5, we demonstrate the relevance of this dataset for MIR research by conducting two case studies on choral intonation assessment and multiple F0-estimation using state-of-the-art algorithms. Finally, in Section 6, we summarize our contributions and experimental results.

2. Prior Work

There is an urgent need for datasets in the field of MIR: annotated data are crucial for training data-driven systems or evaluating methods developed to solve specific tasks. Over the last years, the availability of suitable datasets has triggered research on tasks such as melody extraction (e.g., MedleyDB (Bittner et al., 2014)), music style identification (e.g., Ballroom dataset (Gouyon et al., 2004)), and automatic chord recognition (e.g., Beatles dataset (Harte et al., 2005)).

The datasets closely related to DCS are presented in **Table 1**. Su et al. (2016) created a small dataset for research on choral music. It consists of five short excerpts of Western choral music, ranging from 18 to 40 seconds in length. The dataset contains stereo audio recordings and note event annotations, annotated by a professional pianist. Although small in size, this dataset is relevant for multiple-F0-estimation in complex scenarios where sources are similar, (e.g., voices of a choir), and where several sources produce the same notes (i.e., unisons).

Over the last years, there has been an increasing interest of the MIR community in analyzing world music (Serra, 2014; Panteli, 2018), including traditional singing (van Kranenburg et al., 2019). A conceptually similar dataset to DCS in terms of recording methodology and utilized microphones is a set of multitrack field recordings of three-voice Georgian vocal music (Scherbaum et al., 2019). The dataset includes 216 songs recorded with video cameras, portable stereo recorders as well as multiple close-up microphones attached to each of the singers. Furthermore, the Erkomaishvili Dataset is a publicly available corpus

Table 1: Comparison of polyphonic singing datasets described in Section 2. The reported durations refer to the total recording duration (not counting multiple tracks per recording if available).

Name/Author	Multitrack	Annotations	Publicly Available	# Recordings	Duration (hh:mm:ss)
Su et al. (2016)	No	MIDI	On Request	5 excerpts	00:02:11
Barbershop Quartets ¹⁰	Yes	MIDI	No	22 songs	00:42:10
Bach Chorales ¹¹	Yes	MIDI	No	26 songs	00:58:20
Scherbaum et al. (2019)	Yes	–	On Request	216 songs	06:04:40
Erkomaishvili Dataset (Rosenzweig et al. 2020)	No	Structure, F0, Score, Onsets	Yes	101 songs	07:05:00
Choral Singing Dataset (CSD) (Cuesta et al., 2018)	Yes	MIDI, F0, Notes	Yes	3 songs	00:07:14
Dagstuhl ChoirSet (DCS)	Yes	MIDI, F0, Beats	Yes	2 songs, exercises	00:55:30

based on historic tape recordings of three-voice traditional Georgian songs performed by the former master chanter Artem Erkomaishvili (Rosenzweig et al., 2020). The dataset includes digital sheet music, F0- and onset annotations of the three voices as well as annotations of the overdubbing-based recording structure.

In the context of Western polyphonic vocal music, we find very few multitrack datasets. Two examples are datasets from a commercial application that have been used by Schramm and Benetos (2017); McLeod et al. (2017): the Barbershop Quartets¹⁰ and the Bach Chorales.¹¹ Both datasets contain separate tracks for each of the four SATB singers and an additional track with a stereo mix. The Barbershop recordings comprise 22 songs with a total length of 42 minutes, whereas the Bach Chorales contain 26 recordings with a total length of 58 minutes. The audio recordings and the accompanying synchronized MIDI files are not freely available.

The Choral Singing Dataset (CSD) (Cuesta et al., 2018) is a publicly available dataset of Western polyphonic vocal music.¹² The CSD consists of multitrack recordings of three SATB choral pieces: *Locus Iste* by Anton Bruckner, *Niño Dios d'Amor Herido* by Francisco Guerrero, and *El Rossinyol*, a popular Catalan song, performed by a small choir of 16 singers. The four singers of each choir section were recorded simultaneously in the same room with individual handheld dynamic microphones. However, the different sections were recorded separately where a MIDI track served as reference. The recording length of the three songs is around seven minutes. Furthermore, the CSD includes synchronized MIDI files, note annotations per choir section, and F0-annotations. In summary, the CSD is most similar to our dataset in terms of musical aspects. Further similarities and differences of the CSD to our dataset are discussed in Section 3.3.

3. Dagstuhl ChoirSet

In this section, we describe all components of DCS. In Section 3.1, we give details on the choir settings as well as the recorded pieces and exercises. Then, we explain the recording setup of the multitrack recordings in Section 3.2 and discuss the different dimensions of DCS in

Section 3.3. Subsequently, we elaborate on the manually created beat annotations in Section 3.4. Furthermore, we provide details on the time-aligned score representations in Section 3.5. Finally, we describe the automatically extracted F0-trajectories in Section 3.6.

3.1 Choir Settings and Musical Content

In total, 13 singers (Dagstuhl seminar participants) took part in the recording session. All singers have provided their consent to publish the recorded material for research purposes under a Creative Commons license. The Full Choir consisted of two sopranos, two altos, four tenors, and five basses. From the Full Choir, we selected two soloistic SATB quartets (Quartet A and Quartet B) with four different singers each. The singers had diverse musical backgrounds (from hobby musicians to such holding a music degree) as well as varying levels of experience in (choir) singing within different musical genres. These experiences ranged from singers who had never sung in a choir before to a professional singer with many years of training. Considering that the singers had not sung in this constellation before the Dagstuhl seminar and had only few rehearsals together (3 sessions of roughly 1 hour length), the recorded choir and quartets may be representative of an amateur choir level, with individual skills partly exceeding that level. Rehearsals and recorded performances were also conducted by a Dagstuhl seminar participant, who is a professional composer with solid experience in conducting semi-professional choirs, orchestras, and big bands. We recorded two pieces as well as several intonation exercises with the full choir and the two quartets. The central piece of DCS is Anton Bruckner's *Locus Iste* (WAB 23) in Latin language. **Figure 2** displays the first eleven measures of the piece's score obtained from the Choral Public Domain Library (CPDL).¹³ This small choir piece of approximately three minutes' duration is musically interesting, containing several melodic and harmonic challenges such as chromatic parts and covering a large part of each voice's tessitura (S: B3-G5, A: G3-B4, T: C3-E4, B: F2-C4). Beyond that, the piece is part of the CSD (Cuesta et al., 2018) (see Section 2), thus allowing for interesting

Allegro moderato

1 *p* 2 3 4 *mf* 5 6 7 *f* 8 9 *p* 10 11 12

Lo-cus i-ste: a De-o fa-ctus est, lo-cus i-ste a De-o fa-ctus est, a De-o, De-o fa-ctus est

Lo-cus i-ste: a De-o fa-ctus est, lo-cus i-ste a De-o fa-ctus est, a De-o, De-o fa-ctus est

Lo-cus i-ste: a De-o fa-ctus est, lo-cus i-ste a De-o fa-ctus est, a De-o, De-o fa-ctus est

Lo-cus i-ste: a De-o fa-ctus est, lo-cus i-ste a De-o fa-ctus est, a De-o, De-o fa-ctus est

I II III

Figure 2: Anton Bruckner, *Locus Iste* WAB 23 (measures 1 to 11). The score was obtained from CPDL and edited by Brian Marble.¹³

comparative studies across datasets. Furthermore, we selected the piece *Tebe Poem* by the Bulgarian composer Dobri Hristov.¹⁴ Both pieces are written for SATB choirs in four parts. In addition to these two pieces, the dataset contains a set of vocal exercises of different difficulties and forms taken from the book *Choral Intonation* (Alldahl, 1990). The exercises include scales, long and stable notes, chords, cadences, and a variety of intonation exercises. The additional recordings are potentially interesting to study aspects of ensemble singing such as interval intonation, F0-agreement in unison singing, and intonation drift in a cappella performances.

3.2 Multitrack Recordings

During the recording session, which took place in a Dagstuhl seminar room, we recorded multiple takes of the different pieces and settings. An overview of the recorded material in DCS is presented in **Table 2**. The reported durations refer to the accumulated durations of all takes for a specific piece and setting (not counting multiple tracks per take). The different choir settings were recorded using multiple microphones. In order to record the overall performance, we used an ORTF stereo microphone (Schoeps MSTC 64 U) spaced ca. 3 m away from the singers. The recorded stereo microphone signal is referred to as STM signal in the following. Furthermore, we used several close-up microphones to record individual singers. The recording setup for one singer, which is illustrated in **Figure 3**, includes a handheld dynamic microphone (Sennheiser MD421 II), a headset microphone (DPA 4066F), and a larynx/throat microphone (Albrecht AE 38 S2a). In the following, we abbreviate the three microphone types as DYN, HSM, and LRX respectively.

LRX microphones have shown to be beneficial for analyzing voices of individual singers in polyphonic vocal music (Scherbaum et al., 2015, 2018). Being attached to the skin at the human throat, LRX microphones nicely capture the pitch of the singing voice. Furthermore, compared to other conventional microphones such as DYN microphones, LRX microphones are robust to environmental noise, e.g., the voices of neighbouring singers. However, due to the missing contributions of the vocal tract, LRX signals primarily serve as analysis signals. To illustrate the microphone differences, magnitude spectrograms of LRX and DYN microphone signals for a tenor singer in a quartet setting are shown in **Figure 4a**. The shown excerpts correspond to the marked *Locus Iste* passage in **Figure 2**. It can be observed that the LRX signal is cleaner than the DYN signal. This becomes evident especially in Part II (middle part of the marked passage), where the solo bass voice leaks more strongly into the DYN signal than into the LRX signal of the tenor.

For our recordings, we had four DYN, three HSM and eight LRX microphones available. The complete setup as shown in **Figure 3** could only be used for three singers—other singers were equipped with two, one, or no individual microphone(s). Note that we distributed the microphones such that at least one singer of each part was captured with one LRX and one DYN microphone. The microphone signals were recorded using one RME

Table 2: Overview of the audio recordings in DCS. The third column indicates the number of takes available for each piece and the last column refers to the total duration of all takes together.

Piece	Setting	# Takes	Duration (mm:ss)
<i>Locus Iste</i>	Full Choir	3	07:22
	Quartet A	7	16:26
	Quartet B	6	14:02
<i>Tebe Poem</i>	Full Choir	5	05:27
	Quartet A	2	02:30
Exercises	Full Choir	33	06:00
	Quartet A	25	03:43
Total		81	55:30

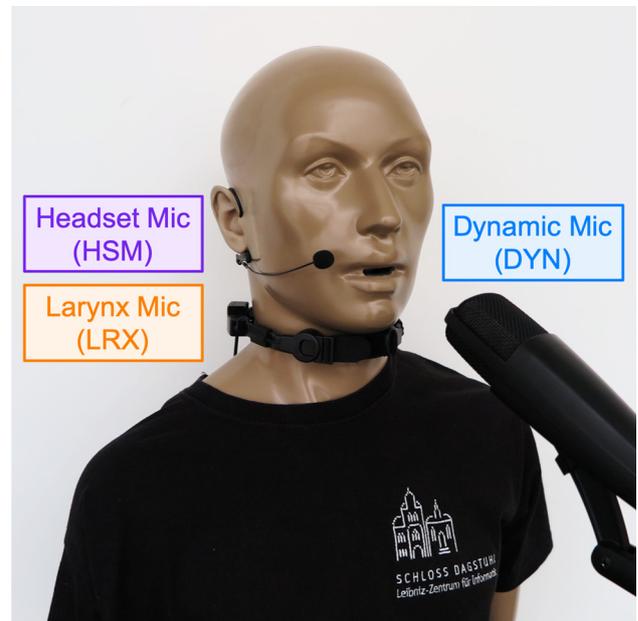


Figure 3: Microphone setup for one singer.

Fireface UFX audio interface, two 8-channel RME Micstasy A/D converters, and the Digital Audio Workstation (DAW) Logic Pro X running on an Apple MacBook Pro (see **Figure 5**). Furthermore, we created an additional reverb version of the stereo microphone signal using the *ChromaVerb* plug-in in Logic Pro X with a decay time of 2 seconds. After recording, all tracks were exported from the DAW and subsequently cut according to manually set cut points using the tool PySox (Bittner et al., 2016). PySox is an open source library that provides a Python interface to SoX (Sound exchange),¹⁵ a command line tool for sound processing. The cut tracks are available in DCS as monophonic WAV files with a sampling rate of 22050 Hz.

3.3 Dagstuhl ChoirSet Dimensions

DCS offers different musical and acoustical dimensions, which are summarized in **Table 3**. We refer to the dimensions as Song, Setting, Take, Voice, and Microphone. The

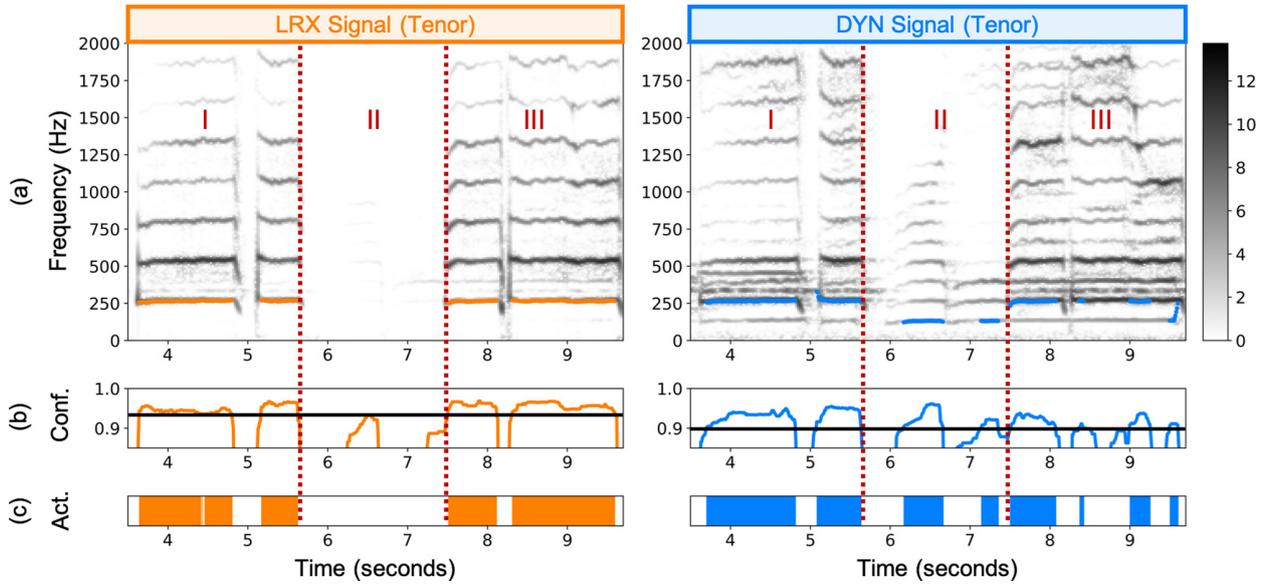


Figure 4: Comparison of LRX and DYN signals from a tenor singer. Excerpts correspond to the marked *Locus Iste* passage in Figure 2. **(a)** Magnitude spectrograms. CREPE F0-trajectories are plotted on top in the respective colors. **(b)** Smoothed CREPE confidence. **(c)** Binarized trajectory activations obtained by thresholding smoothed confidence (LRX threshold: 0.935, DYN threshold: 0.9).



Figure 5: Screenshot (detail) of digital audio workstation (Logic Pro X) with multiple tracks.

Song dimension consists of the two choral pieces *Locus Iste* and *Tebe Poem* as well as the systematic exercises. The Setting dimension includes the three choir settings: Full Choir, Quartet A, and Quartet B. The Take dimension indicates the number of takes. The Voice dimension is defined by the singers present in the signal—either one of the SATB sections or the mixture of all sections recorded by the STM microphone. Finally, the Microphone dimension refers to the microphone types used to record the singers.

The multiple dimensions of DCS make it unique when compared to related datasets such as the CSD (Cuesta et al., 2018). The main differences between the CSD and DCS lie in the Setting, Take, and Microphone dimensions. The CSD includes one singer setting, a single take per song and one

Table 3: DCS dimensions.

Dimension	Shortcut	Meaning
Song	LI	<i>Locus Iste</i>
	TP	<i>Tebe Poem</i>
	SE	Systematic Exercises
Setting	FullChoir	Full Choir Setting
	QuartetA	Quartet A Setting
	QuartetB	Quartet B Setting
Take	Take	Take Number
Voice	S	Soprano
	A	Alto
	T	Tenor
	B	Bass
Microphone	Stereo	Stereo Mic
	StereoReverb	Stereo Mic Reverb
	LRX	Larynx Mic
	DYN	Dynamic Mic
	HSM	Headset Mic
	STR	Stereo Mic R
	STL	Stereo Mic L
STM	Stereo Mic L+R	

microphone type. Furthermore, the CSD choir sections were recorded separately, while all singers were captured at the same time in DCS. The different recording setup in DCS enables studies on interactions between sections. However, as opposed to the Full Choir setting in DCS, the recorded choir in the CSD is larger and balanced in the

number of singers per section. Therefore, CSD allows for more detailed studies on singer interaction within choir sections.

In order to account for the variety of different dimensions, we developed a filename convention for all audio and annotation files included in DCS. The general format of the filenames is the following (cf. **Table 3**): `DCS_{Song}_{Setting}_Take{#}_{Voice}{#}_{Microphone}.{Suffix}`. For example, `DCS_LI_FullChoir_Take02_T2_LRX.wav` refers to the audio signal from the larynx microphone (LRX) of the second tenor (T2) in the Full Choir setting (FullChoir) during the second take (Take02) of *Locus Iste* (LI). Note that the files with microphone shortcut STM contain a mono mix of the left and right channel of the stereo microphone.

3.4 Manual Beat Annotations

The beat is a key unit of the temporal structure of music (Goto and Muraoka, 1997). As stated by Robertson (2012), when beat annotations are manually generated by tapping along to an audio signal, they reflect the ability of the annotator to *produce* the beats rather than their *perception*. In such cases, the *produced* beat annotations can be subsequently refined by iteratively listening and modifying them according to perceptual cues. Following this premise, we generated beat annotations for all STM signals of *Locus Iste* and *Tebe Poem* in a two-stage process: in the first stage, annotations were manually created by an annotator with some musical background. The *annotation by tapping* feature in Sonic Visualiser (Cannam et al., 2010b) was used for this task. Sonic Visualiser is an open source software for generating manual annotations of various kinds. In the second stage, annotations were reviewed and refined by a second, experienced annotator using the same software.

These beat annotations are provided as comma-separated value (CSV) files with two columns. The first column contains timestamps in seconds, whereas the second column contains beat and measure information provided as floating point numbers to three decimal places. The part in front of the decimal point encodes the measure number. The part after the decimal point indicates the beat position inside the measure. For example, in 4/4 time, each beat is represented as an increment of $1/4 = 0.250$, and therefore the beat positions are given as 1.000, 1.250, 1.500, 1.750, 2.000, 2.250, 2.500....

3.5 Time-Aligned Score Representations

In order to obtain a musical reference for the different performances of *Locus Iste* and *Tebe Poem*, we aligned MIDI representations of the pieces to the STM signals using the beat annotations from Section 3.4. The MIDI files were obtained from the CPDL (see Section 3.1). For synchronization, we used the dynamic time warping pipeline from Ewert et al. (2009) and Müeller et al. (2004) that uses the beat annotations as anchor points for the alignment. In order to facilitate data parsing and processing, we converted the aligned MIDI files to CSV files using `pretty_midi` (Raffel and Ellis, 2014), a Python

library for processing and converting MIDI files. For each STM signal, DCS contains one separate CSV file per section (as opposed to MIDI files that include all sections). Each CSV file contains three columns, which represent note onset in seconds, note offset in seconds, and MIDI pitch. The number of rows is equal to the number of notes in the piece.

3.6 Fundamental Frequency Trajectories

One of the most important cues for computational studies on choral singing and choral intonation are the F0-trajectories of the individual singers' voices (Cuesta et al., 2018; Dai and Dixon, 2017, 2019). However, annotating F0-trajectories from polyphonic mixtures is cumbersome and requires a lot of labor-intensive work. We exploit the multitrack nature of DCS to automatically compute the F0-trajectories of each singer from the close-up microphone signals using two state-of-the-art algorithms for monophonic F0-estimation: pYIN (Mauch and Dixon, 2014) and CREPE (Kim et al., 2018).

The pYIN annotations were obtained using the pYIN Vamp Plug-in¹⁶ for Sonic Annotator (Cannam et al., 2010a). For pYIN, we used an FFT size of 2048 and a hop size of 221 samples, which corresponds to around 10 ms for a sampling rate of 22050 Hz. We used the algorithm in the `smoothedpitchtrack` mode, which uses a hidden Markov model (HMM) and Viterbi decoding to smooth the F0-estimates. In addition, we configured the plugin to output negative F0-values in frames that are estimated as unvoiced (`outputunvoiced=2`) as well as the probability of each frame to be voiced (`output=voicedprob`). For CREPE, we used the CREPE Python package¹⁷ with the model capacity set to `full`, Viterbi smoothing activated, a default hop size of 10 ms, and a default input size of 1024 samples. Similar hop sizes were used with both methods for an easier comparison. The F0-trajectories are stored in CSV files with three columns. The first two columns contain the timestamps in seconds and the F0-values in Hz. In the case of pYIN, the third column contains the probabilities of the frames to be voiced. In the case of CREPE, the third column contains the confidence as provided by the algorithm. The confidence is a number between 0 and 1 that indicates the reliability of an F0-estimate.

In order to validate the automatically extracted F0-trajectories, we generated manual F0-annotations for all voices of two quartet recordings based on the LRX signals. The annotations were made by a sound engineer with over ten years' training on saxophone using the tool Tony (Mauch et al., 2015) and are included in DCS as CSV files. For evaluation, we use common evaluation metrics for melody extraction as detailed by Poliner et al. (2007); Salamon et al. (2014). The metrics Voicing Recall (VR) and Voicing False Alarm (VFA) measure the accuracy of the algorithm's voice activity estimation. The metrics Raw Pitch Accuracy (RPA) and Raw Chroma Accuracy (RCA) measure the proportion of frames for which the estimated F0-trajectory lies within 50 cents (half a semitone) of the reference (RCA ignores octave errors). Additionally, the Overall Accuracy (OA) is a combined metric that accounts

for both voice activity and F0-accuracy. We use the open source toolbox `mir_eval` (Raffel et al., 2014) to compute the evaluation metrics. In our experiments, we derive the voice activity for F0-trajectories extracted by CREPE by choosing a confidence threshold that maximizes the overall accuracy. The evaluation results averaged over the two recordings for pYIN and CREPE (8 LRX, 6 HSM, 8 DYN trajectories per algorithm) are given in **Tables 4** and **5**, respectively. The standard deviations are given in brackets. Both algorithms perform most accurately on the LRX signals (0.93 of overall accuracy), slightly less accurate on DYN signals and least accurate on HSM signals. This is expected, since the F0 of the voice is more dominant in LRX signals than in DYN or HSM signals (see Section 3.2). The overall performance of both algorithms is similar on LRX and DYN signals and deviates for HSM signals, where CREPE performs better than pYIN.

In the following, we further analyze the differences between the microphone signals. **Figure 4** illustrates the F0-trajectories from a tenor singer extracted from LRX and DYN signals using CREPE. The CREPE confidence values are depicted in **Figure 4b**. For visualization purposes, the confidences are smoothed with a median filter of length 210 ms. Thresholding the smoothed confidence values with a threshold of 0.935 for the LRX confidence and a threshold of 0.9 for the DYN confidence leads to the binary activations depicted in **Figure 4c** and the F0-trajectories depicted in **Figure 4a**. Note that the thresholds are chosen exemplarily to show the differences between the microphones. In Part I, CREPE shows similar confidence values for both microphone signals when the tenor is singing. Part II shows significant differences between the two microphones. In this part, low confidence values are expected since the tenor is not active. Still, CREPE shows some confidence for both microphone signals due to cross-talk of the bass voice. However, one can find a suitable threshold for the LRX confidence to avoid an F0-output. Since the cross-talk is much stronger in the DYN signal, there exists no meaningful threshold that suppresses any F0-output in Part II of the DYN signal. In Part III, the F0-trajectory of the DYN microphone suffers from confusions with the bass voice even though the tenor is singing.

4. Dagstuhl ChoirSet Interfaces

The main goal of our work is to create a freely available and easy-to-access dataset in order to support MIR research on a cappella choral music. To this end, we provide several interfaces to interact with the dataset. As the most important step, we make the dataset publicly available in order to support scientific exchange and ensure reproducibility of scientific results. We decided to host DCS on Zenodo,⁷ an Open Science platform, which supports sharing and distributing scientific data. As main features, the platform provides versioning and citeable Digital Object Identifiers (DOIs) for uploaded data.

However, Zenodo is a data repository and does not offer to play back the audio files in the browser. The interdisciplinary field of MIR benefits from interfaces that help to lower access barriers to datasets by providing direct, intuitive, and comprehensive access. This can be accomplished by means of interactive interfaces, e.g., with playback functionalities (Gasser et al., 2015; Jeong et al., 2017; Röwenstrunk et al., 2015). As one contribution, we created a publicly accessible web-based interface,⁸ which hosts the multitrack audio data. The entry page of the interface is subdivided into a “Music Recordings” section providing links to the *Locus Iste* and *Tebe Poem* recordings as well as a “Systematic Exercises and Additional Recordings” section. Furthermore, the interface allows for searching and sorting of specific recordings. Each multitrack recording has an individual sub-page with an open source audio player (Werner et al., 2017) with score-following functionality (Zalkow et al., 2018) that allows for seamless switching between the different tracks.

Accompanying dataset-specific processing tools simplify the usage of datasets (Bittner et al., 2014, 2019). Therefore, we created a Python toolbox named `DCStoolbox`⁹ that accompanies the release of the dataset. The toolbox provides basic functions to parse and load data from DCS, which are demonstrated in a Jupyter notebook. Additionally, the toolbox includes scripts to reproduce the computed F0-trajectories from Section 3.6 and an Anaconda¹⁸ environment file that specifies all Python packages required to run the toolbox functions.

Table 4: Evaluation results for pYIN trajectories averaged over two quartet recordings.

Mic	VR	VFA	RPA	RCA	OA
LRX	0.99 (0.00)	0.11 (0.06)	0.95 (0.02)	0.95 (0.01)	0.93 (0.03)
HSM	0.98 (0.01)	0.33 (0.09)	0.81 (0.10)	0.91 (0.04)	0.77 (0.08)
DYN	0.99 (0.00)	0.16 (0.11)	0.93 (0.04)	0.95 (0.01)	0.90 (0.05)

Table 5: Evaluation results for CREPE trajectories averaged over two quartet recordings.

Mic	VR	VFA	RPA	RCA	OA
LRX	0.96 (0.01)	0.12 (0.02)	0.96 (0.01)	0.96 (0.01)	0.93 (0.02)
HSM	0.92 (0.02)	0.32 (0.08)	0.91 (0.01)	0.91 (0.02)	0.84 (0.02)
DYN	0.93 (0.01)	0.18 (0.07)	0.93 (0.01)	0.93 (0.01)	0.90 (0.02)

5. Applications to MIR Research

In this section, we demonstrate the potential of DCS for MIR research by means of two case studies. In the first case study discussed in Section 5.1, the goal is to evaluate and compare the intonation quality of quartet performances using a recently published intonation measure (Weiß et al., 2019). In the second case study, conducted in Section 5.2, we consider the task of multiple F0-estimation. More specifically, we apply a state-of-the-art approach (Bittner et al., 2017) on different recordings and show the benefits of our multitrack recordings for multiple-F0-estimation in polyphonic vocal music.

5.1 Intonation Quality of Quartet Performances

A central challenge for a cappella singers is the adjustment of pitch in order to stay in tune relative to the fellow singers. Even if choirs achieve good local intonation, they may suffer from intonation drifts slowly evolving over time (Devaney, 2011). Algorithms that attempt to measure intonation quality have to account for such intonation drifts. A recently published approach measures the distance between the recording's local salient frequency content and a shifted 12-tone equal-tempered (12-TET) grid (Weiß et al., 2019). Although choirs often aim for just intonation, the 12-TET scale has been used to approximate intonation in Western choral performances (Gnann et al., 2011). The intonation measure requires as input the F0s and harmonic partials (integer multiples of the F0) together with their respective amplitudes for the four singing voices. In a frame-wise fashion, a grid-shift parameter is computed that minimizes the distance between the F0s partials and the shifted 12-TET grid. As output, the approach returns a frame-wise intonation cost (IC) that reflects the remaining distance from the optimally shifted 12-TET grid. The IC is bounded in the interval $[0, 1]$, where small values indicate good local intonation, and large values indicate local intonation deviations. In the following, we use this approach to compare the performances of Quartet A and B in our DCS.

Weiß et al. (2019) show that multitrack recordings of the individual voices are beneficial for estimating the frequency and amplitude information required to compute the IC. For our case study, we make use of the recorded LRX and DYN signals as follows. We obtain the frequency information from the extracted pYIN F0-trajectories of the LRX signals (see Section 3.6). Using the time-aligned score representations from Section 3.5, we restrict the trajectories to regions where the respective voices are active. We obtain the amplitude information from a magnitude spectrogram representation of the DYN signals at the locations of the extracted LRX F0-trajectories and their harmonic partials. In our experiments, we consider 16 harmonic partials. Subsequently, we compute IC measure curves for all quartet recordings of *Locus Iste* in DCS. In order to compare the different takes, we map the curves on a common time axis in measures using the measure information encoded in the beat annotations from Section 3.4.

The averaged IC curves for six recordings of Quartet A and five recordings of Quartet B are depicted in **Figure 6**. To remove local outliers, we post-process the IC curves using a moving median filter of length 21 frames. Note that the IC is zero for silent regions and small for monophonic passages where only one singer is active (see measures 12, 20/21, and 43). Overall, the curves exhibit a similar progression. For both curves, we observe higher IC values in the passage from measures 13 to 20. This passage is challenging to sing due to the highly chromatic voice leading and the jumps in the bass part. Furthermore, the passage from measure 40 to 42 exhibits higher intonation costs for both quartets—a passage which is highly chromatic. The largest differences between the quartets can be found in the last part of the piece (measures 44 to 48). For this passage, Quartet B achieves a better intonation quality on average than Quartet A, especially in the intonation of the final chord of the piece.

This short case study indicates the potential of our recordings for studying intonation in polyphonic a cappella music. Furthermore, our data can form a starting

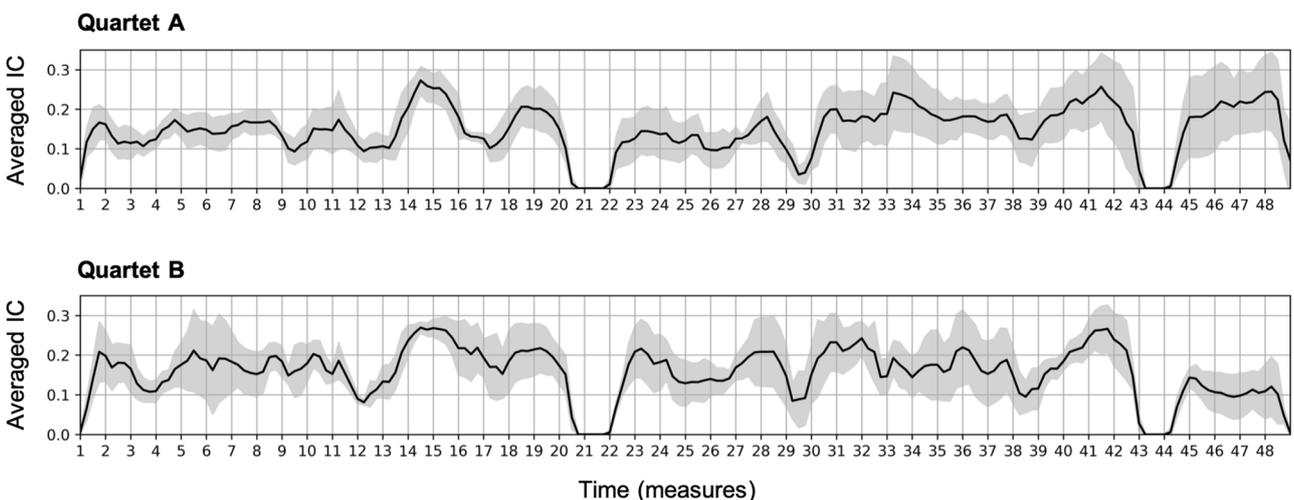


Figure 6: Averaged intonation cost (IC) measures for six takes of *Locus Iste* by Quartet A and five takes by Quartet B. The local standard deviations are indicated in light grey.

point for future studies on singer interaction in amateur choirs.

5.2 Multiple-F0-Estimation in A Cappella Singing

Multiple-F0-estimation is defined as the task of estimating the F0s of several concurrent sounds in a polyphonic signal (Klapuri, 2008, 2006). This task is particularly challenging for polyphonic vocal music (Schramm et al., 2017; Su et al., 2016). In a cappella choral singing, we find multiple singers with similar timbres singing in harmony, thus producing overlapping harmonics (Cuesta et al., 2019). Furthermore, it is very common that several singers sing the same part (unison), but produce slightly different frequencies. However, MIR research on multiple-F0-estimation in polyphonic vocal music has so far been focusing on SATB quartets and there exist no suitable methods for multiple-F0-estimation in larger ensembles with multiple singers per part. The Full Choir recordings in DCS constitute a starting point for further research in this direction.

In the following, we show the potential of DCS by applying a state-of-the-art multiple-F0-estimation algorithm on different scenarios offered by the DCS quartet recordings. The first scenario consists of applying the algorithm on a mix of all DYN signals. In the second and third scenario, the algorithm is applied on the STM signal (room microphone) with and without additional reverb. In particular, we consider the recordings of *Locus Iste* from Quartet A (Take 3).

In our case study, we use the *DeepSalience* method (Bittner et al., 2017), a deep convolutional neural network trained to produce a pitch salience representation (enhanced time–frequency representation) of the input signal, which contains values in the range $[0, 1]$. This salience representation is thresholded such that only time–frequency bins with a salience value above the chosen threshold remain. These remaining bins correspond to the multiple-F0-estimates. Although the model is not specifically trained for polyphonic vocal music, it was found to obtain the best performance for multiple-F0-estimation

in vocal quartets (Cuesta et al., 2019). For the evaluation, we exploit the multitrack nature of DCS. In particular, we take the previously extracted pYIN F0-trajectories from the LRX signals as reference (see Section 3.6). Note that these trajectories are the output of an algorithm. Although our evaluation reveals they are very accurate (see **Table 4**), they still contain some errors. As evaluation metrics, we use the standard multiple-F0-estimation metrics Precision, Recall, and F-Score. For a detailed description of these metrics, we refer to Bittner (2018, Chapter II, Section 6.3). The evaluation metrics were computed using the `mir_eval` library (Raffel et al., 2014).

We experimented with several thresholds between 0.05 and 0.5, and found 0.1 to obtain the best results on the studied quartet recordings with respect to our evaluation metrics. However, instead of comparing absolute values (which is problematic for automatically extracted reference F0-trajectories), we want to focus on relative differences between the different scenarios. **Figure 7a** shows excerpts of the computed multiple-F0-estimates for the mix of DYN signals and the STM signal with reverb obtained by thresholding the salience representations with a threshold of 0.1. **Figure 7b** shows the evaluation results for all three scenarios. From the F-Score values, we observe that the algorithm performs best for the DYN signal mix of Quartet A. Furthermore, we observe that an increasing amount of reverb in the recordings goes along with a decreasing overall performance of the algorithm. This indicates that reverb further complicates the task of multiple-F0-estimation. The Precision and Recall measures give further insights into this observation. While Precision is lower in the scenario with reverb, Recall is not affected. In reverb conditions, sung notes become temporally smeared, leading to a temporal mismatch between the reference F0-trajectories from the LRX signals and the audio recording. For this reason, the number of false positives increases, causing Precision to decrease. This effect can be seen by comparing the red marked areas in **Figure 7a**. We leave a more detailed analysis of these effects to future studies.

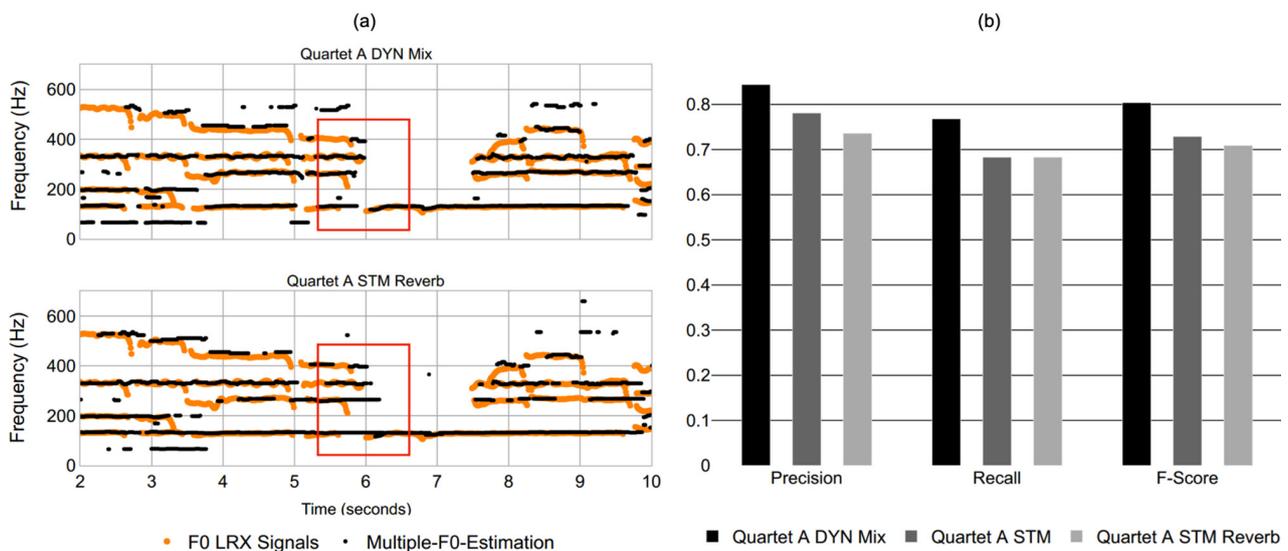


Figure 7: Multiple-F0-estimation using *DeepSalience* (Bittner et al., 2017) with a threshold of 0.1. **(a)** Estimation results (excerpts) for the mix of DYN signals and the STM signal with reverb. **(b)** Evaluation metrics for all scenarios.

In summary, this brief case study indicates that the DCS is a versatile and challenging resource to develop and test algorithms for multiple-F0-estimation in polyphonic a cappella vocal music. Furthermore, the time-aligned score representations could serve as a reference for the evaluation of note-tracking algorithms. This requires accounting for intonation drifts of the choirs, which can, e.g., be determined from the F0-annotations.

6. Conclusions

In this paper, we presented Dagstuhl ChoirSet—a publicly accessible multitrack dataset of a cappella choral music for MIR research. This work is based on our recordings of an amateur vocal ensemble we gathered at an MIR seminar at Schloss Dagstuhl. As main feature of the dataset, the singers were recorded using different close-up microphones including dynamic, headset, and larynx microphones. As part of our work, we curated the recorded material and manually generated beat annotations as well as time-aligned sheet music representations. Furthermore, we automatically extracted F0-trajectories for all close-up microphone tracks. The dataset is released together with an interactive web-based interface and a Python toolbox to provide convenient access. In summary, the different musical and acoustical dimensions of DCS open up a variety of new and challenging scenarios for MIR research.

Notes

- ¹ <https://europeanchoralassociation.org>.
- ² <https://www.chorusamerica.org>.
- ³ <https://www.carus-verlag.com/en/digital-media/carus-music-the-choir-app>.
- ⁴ <https://www.singerhood.com>.
- ⁵ <https://trompamusic.eu/choir-singers>.
- ⁶ <https://www.dagstuhl.de/19052>.
- ⁷ <https://doi.org/10.5281/zenodo.3956666>.
- ⁸ <https://www.audiolabs-erlangen.de/resources/MIR/2020-DagstuhlChoirSet>.
- ⁹ <https://github.com/helenacuesta/ChoirSet-Toolbox>.
- ¹⁰ <https://www.pgmusic.com/barbershopquartet.htm>.
- ¹¹ <https://www.pgmusic.com/bachchorales.htm>.
- ¹² <https://zenodo.org/record/2649950>.
- ¹³ http://www1.cpd1.org/wiki/images/9/94/Locus_Iste_rev.pdf.
- ¹⁴ [http://www3.cpd1.org/wiki/index.php/Tebe_Poem_\(Dobri_Hristov\)](http://www3.cpd1.org/wiki/index.php/Tebe_Poem_(Dobri_Hristov)).
- ¹⁵ <http://sox.sourceforge.net/>.
- ¹⁶ <https://code.soundsoftware.ac.uk/projects/pyin>.
- ¹⁷ <https://github.com/marl/crepe>.
- ¹⁸ <https://www.anaconda.com/distribution/>.

Acknowledgements

We would like to thank all participants of the Dagstuhl seminar on “Computational Methods for Melody and Voice Processing in Music Recordings” for contributing to the recordings. Furthermore, we thank Schloss Dagstuhl for hosting the seminar and for supporting the recording session. We thank Vlora Arifi-Müller, Simon Schwär and Moritz Berendes for helping with the annotations and

Florian Schuberth for helping with Python implementations. Sebastian Rosenzweig is supported by the German Research Foundation (DFG MU 2686/12-1). Helena Cuesta is supported by FI Predoctoral Grant from AGAUR (Generalitat de Catalunya) and by the European Commission under the TROMPA project (H2020 770376). The International Audio Laboratories Erlangen are a joint institution of the Friedrich-Alexander-Universität Erlangen-Nürnberg (FAU) and Fraunhofer Institute for Integrated Circuits IIS.

Competing Interests

The authors have no competing interests to declare.

Author Contributions

Sebastian Rosenzweig and Helena Cuesta substantially contributed to the recording process, dataset creation and writing of this article. Christof Weiß took over the musical coordination as well as the conducting during rehearsals and recordings. Frank Scherbaum contributed to the technical coordination of the recording session and provided his microphones. Emilia Gómez and Meinard Müller supervised this work and contributed to dataset creation and writing of the article. All authors have read and agreed to the published version of the manuscript.

References

- Alldahl, P.-G.** (1990). *Choral Intonation*. Gehrmans Musikförlag.
- Benetos, E., Dixon, S., Duan, Z., & Ewert, S.** (2019). Automatic music transcription: An overview. *IEEE Signal Processing Magazine*, 36(1), 20–30. DOI: <https://doi.org/10.1109/MSP.2018.2869928>
- Bittner, R. M.** (2018). *Data-driven fundamental frequency estimation*. PhD thesis, New York University.
- Bittner, R. M., Fuentes, M., Rubinstein, D., Jansson, A., Choi, K., & Kell, T.** (2019). mirdata: Software for reproducible usage of datasets. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 99–106. Delft, The Netherlands.
- Bittner, R. M., Humphrey, E., & Bello, J. P.** (2016). PySox: Leveraging the audio signal processing power of SoX in Python. In *Late Breaking and Demo Papers, International Society for Music Information Retrieval Conference (ISMIR) Conference*.
- Bittner, R. M., McFee, B., Salamon, J., Li, P., & Bello, J. P.** (2017). Deep salience representations for F0 tracking in polyphonic music. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 63–70. Suzhou, China.
- Bittner, R. M., Salamon, J., Tierney, M., Mauch, M., Cannam, C., & Bello, J. P.** (2014). MedleyDB: A multitrack dataset for annotation-intensive MIR research. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 155–160. Taipei, Taiwan.
- Blaauw, M., & Bonada, J.** (2017). A neural parametric singing synthesizer modeling timbre and expression from natural songs. *Applied Sciences*, 7(1313). DOI: <https://doi.org/10.3390/app7121313>

- Böck, S., Davies, M. E. P., & Knees, P.** (2019). Multitask learning of tempo and beat: Learning one to improve the other. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 486–493. Delft, The Netherlands.
- Cannam, C., Jewell, M. O., Rhodes, C., Sandler, M., & d'Inverno, M.** (2010a). Linked data and you: Bringing music research software into the semantic web. *Journal of New Music Research*, 39(4), 313–325. DOI: <https://doi.org/10.1080/09298215.2010.522715>
- Cannam, C., Landone, C., & Sandler, M. B.** (2010b). Sonic Visualiser: An open source application for viewing, analysing, and annotating music audio files. In *Proceedings of the International Conference on Multimedia*, pages 1467–1468. Florence, Italy. DOI: <https://doi.org/10.1145/1873951.1874248>
- Cano, E., FitzGerald, D., Liutkus, A., Plumbley, M. D., & Stöter, F.** (2019). Musical source separation: An introduction. *IEEE Signal Processing Magazine*, 36(1), 31–40. DOI: <https://doi.org/10.1109/MSP.2018.2874719>
- Cano, E., Schuller, G., & Dittmar, C.** (2014). Pitchinformed solo and accompaniment separation towards its use in music education applications. *EURASIP Journal on Advances in Signal Processing*, 2014(23). DOI: <https://doi.org/10.1186/1687-6180-2014-23>
- Chandna, P., Blaauw, M., Bonada, J., & Gómez, E.** (2019). A vocoder based method for singing voice extraction. In *Proceedings of the 44th IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Brighton, UK. IEEE. DOI: <https://doi.org/10.1109/ICASSP.2019.8683323>
- Cuesta, H., Gómez, E., & Chandna, P.** (2019). A framework for multi-f0 modeling in SATB choir recordings. In *Proceedings of the Sound and Music Computing (SMC) Conference*, pages 447–453.
- Cuesta, H., Gómez, E., Martorell, A., & Loáiciga, F.** (2018). Analysis of intonation in unison choir singing. In *Proceedings of the International Conference of Music Perception and Cognition (ICMPC)*, pages 125–130. Graz, Austria.
- Dai, J., & Dixon, S.** (2017). Analysis of interactive intonation in unaccompanied SATB ensembles. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 599–605. Suzhou, China.
- Dai, J., & Dixon, S.** (2019). Singing together: Pitch accuracy and interaction in unaccompanied unison and duet singing. *The Journal of the Acoustical Society of America*, 145(2), 663–675. DOI: <https://doi.org/10.1121/1.5087817>
- Devaney, J.** (2011). *An Empirical Study of the Influence of Musical Context on Intonation Practices in Solo Singers and SATB Ensembles*. PhD thesis, McGill University, Montreal, Canada.
- Devaney, J., & Ellis, D. P. W.** (2008). An empirical approach to studying intonation tendencies in polyphonic vocal performances. *Journal of Interdisciplinary Music Studies*, 2(1&2), 141–156.
- Ewert, S., Müller, M., & Grosche, P.** (2009). High resolution audio synchronization using chroma onset features. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pages 1869–1872. Taipei, Taiwan. DOI: <https://doi.org/10.1109/ICASSP.2009.4959972>
- Gasser, M., Arzt, A., Gadermaier, T., Grachten, M., & Widmer, G.** (2015). Classical music on the web – user interfaces and data representations. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 571–577. Málaga, Spain.
- Gnann, V., Kitzka, M., Becker, J., & Spiertz, M.** (2011). Least-squares local tuning frequency estimation for choir music. In *Proceedings of the Audio Engineering Society (AES) Convention*, New York City, USA.
- Gómez, E., Gkiokas, A., Liem, C., Samiotis, I. P., Gutierrez, N., Santos, P., Crawford, T., Weigl, D. M., Goebel, W., Tilburg, M., Sarasua, Á., & Freiburg, B.** (2020). Towards richer online public-domain archives of classical music. *Submitted to Human Computation Journal*.
- Goto, M., & Muraoka, Y.** (1997). Issues in evaluating beat tracking systems. In *Working Notes of the IJCAI-97 Workshop on Issues in AI and Music- Evaluation and Assessment*, pages 9–16.
- Gouyon, F., Dixon, S., Pampalk, E., & Widmer, G.** (2004). Evaluating rhythmic descriptors for musical genre classification. In *Proceedings of the Audio Engineering Society (AES) International Conference*, London, UK.
- Harte, C., Sandler, M. B., Abdallah, S., & Gómez, E.** (2005). Symbolic representation of musical chords: A proposed syntax for text annotations. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 66–71. London, UK.
- Howard, D. M.** (2007). Intonation drift in a capella soprano, alto, tenor, bass quartet singing with key modulation. *Journal of Voice*, 21(3), 300–315. DOI: <https://doi.org/10.1016/j.jvoice.2005.12.005>
- Howard, D. M., Daffern, H., & Breerton, J.** (2013). Four-part choral synthesis system for investigating intonation in a cappella choral singing. *Logopedics Phoniatrics Vocology*, 38(3), 135–142. DOI: <https://doi.org/10.3109/14015439.2013.812143>
- Jeong, D., Kwon, T., Park, C., & Nam, J.** (2017). PerformScore: Toward performance visualization with the score on the web browser. In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Suzhou, China.
- Kim, J. W., Salamon, J., Li, P., & Bello, J. P.** (2018). Crepe: A convolutional representation for pitch estimation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 161–165. Calgary, Canada. DOI: <https://doi.org/10.1109/ICASSP.2018.8461329>
- Klapuri, A. P.** (2006). Multiple fundamental frequency estimation by summing harmonic amplitudes. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 216–221.
- Klapuri, A. P.** (2008). Multipitch analysis of polyphonic music and speech signals using an auditory model. *IEEE Transactions on Audio, Speech, and Language*

- Processing*, 16(2), 255–266. DOI: <https://doi.org/10.1109/TASL.2007.908129>
- Mauch, M., Cannam, C., Bittner, R., Fazekas, G., Salamon, J., Dai, J., Bello, J., & Dixon, S.** (2015). Computer-aided melody note transcription using the Tony software: Accuracy and efficiency. In *Proceedings of the International Conference on Technologies for Music Notation and Representation*.
- Mauch, M., & Dixon, S.** (2014). pYIN: A fundamental frequency estimator using probabilistic threshold distributions. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 659–663. Florence, Italy. DOI: <https://doi.org/10.1109/ICASSP.2014.6853678>
- Mauch, M., Frieler, K., & Dixon, S.** (2014). Intonation in unaccompanied singing: Accuracy, drift, and a model of reference pitch memory. *Journal of the Acoustical Society of America*, 136(1), 401–411. DOI: <https://doi.org/10.1121/1.4881915>
- McLeod, A., Schramm, R., Steedman, M., & Benetos, E.** (2017). Automatic transcription of polyphonic vocal music. *Applied Sciences*, 7(12). DOI: <https://doi.org/10.3390/app7121285>
- Müller, M., Gómez, E., & Yang, Y.** (2019). Computational methods for melody and voice processing in music recordings (Dagstuhl seminar 19052). *Dagstuhl Reports*, 9(1), 125–177.
- Müller, M., Kurth, F., & Röder, T.** (2004). Towards an efficient algorithm for automatic score-to-audio synchronization. In *Proceedings of the International Conference on Music Information Retrieval (ISMIR)*, pages 365–372. Barcelona, Spain.
- Panteli, M.** (2018). *Computational analysis of world music corpora*. PhD thesis, Queen Mary University of London, UK.
- Poliner, G. E., Ellis, D. P., Ehmann, A. F., Gómez, E., Streich, S., & Ong, B.** (2007). Melody transcription from music audio: Approaches and evaluation. *IEEE Transactions on Audio, Speech, and Language Processing*, 15(4), 1247–1256. DOI: <https://doi.org/10.1109/TASL.2006.889797>
- Raffel, C., & Ellis, D. P. W.** (2014). Intuitive analysis, creation and manipulation of MIDI data with pretty_midi. In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Taipei, Taiwan.
- Raffel, C., McFee, B., Humphrey, E. J., Salamon, J., Nieto, O., Liang, D., & Ellis, D. P. W.** (2014). MIR_EVAL: A transparent implementation of common MIR metrics. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 367–372. Taipei, Taiwan.
- Robertson, A.** (2012). Decoding tempo and timing variations in music recordings from beat annotations. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 475–480.
- Rosenzweig, S., Scherbaum, F., Shugliashvili, D., Arifi-Müller, V., & Müller, M.** (2020). Erkomaishvili Dataset: A curated corpus of traditional Georgian vocal music for computational musicology. *Transactions of the International Society for Music Information Retrieval (TISMIR)*, 3(1), 31–41. DOI: <https://doi.org/10.5334/tismir.44>
- Rößenstrunk, D., Prätzlich, T., Betzwieser, T., Müller, M., Szwillus, G., & Veit, J.** (2015). Das Gesamtkunstwerk Oper aus Datensicht – Aspekte des Umgangs mit einer heterogenen Datenlage im BMBF-Projekt “Freischütz Digital”. *Datenbank-Spektrum*, 15(1), 65–72. DOI: <https://doi.org/10.1007/s13222-015-0179-0>
- Salamon, J., Gómez, E., Ellis, D. P. W., & Richard, G.** (2014). Melody extraction from polyphonic music signals: Approaches, applications, and challenges. *IEEE Signal Processing Magazine*, 31(2), 118–134. DOI: <https://doi.org/10.1109/MSP.2013.2271648>
- Scherbaum, F., Loos, W., Kane, F., & Vollmer, D.** (2015). Body vibrations as source of information for the analysis of polyphonic vocal music. In *Proceedings of the International Workshop on Folk Music Analysis*, pages 89–93. Paris, France.
- Scherbaum, F., Mzhavanadze, N., Rosenzweig, S., & Müller, M.** (2019). Multi-media recordings of traditional Georgian vocal music for computational analysis. In *Proceedings of the International Workshop on Folk Music Analysis*, pages 1–6. Birmingham, UK.
- Scherbaum, F., Rosenzweig, S., Müller, M., Vollmer, D., & Mzhavanadze, N.** (2018). Throat microphones for vocal music analysis. In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France.
- Schramm, R., & Benetos, E.** (2017). Automatic transcription of a cappella recordings from multiple singers. In *AES International Conference on Semantic Audio*. Audio Engineering Society.
- Schramm, R., McLeod, A., Steedman, M., & Benetos, E.** (2017). Multi-pitch detection and voice assignment for a cappella recordings of multiple singers. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 552–559. Suzhou, China.
- Serra, X.** (2014). Creating research corpora for the computational study of music: The case of the CompMusic project. In *Proceedings of the AES International Conference on Semantic Audio*, London, UK.
- Su, L., Chuang, T.-Y., & Yang, Y.-H.** (2016). Exploiting frequency, periodicity and harmonicity using advanced time-frequency concentration techniques for multipitch estimation of choir and symphony. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 393–399. New York City, USA.
- Sundberg, J.** (1987). *The Science of the Singing Voice*. Northern Illinois University Press.
- Thomas, V., Fremerey, C., Müller, M., & Clausen, M.** (2012). Linking sheet music and audio – challenges and new approaches. In Müller, M., Goto, M., & Schedl, M., Editors, *Multimodal Music Processing*, volume 3 of *Dagstuhl Follow-Ups*, pages 1–22. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, Dagstuhl, Germany.

- van Kranenburg, P., de Bruin, M., & Volk, A.** (2019). Documenting a song culture: The Dutch Song Database as a resource for musicological research. *International Journal on Digital Libraries*, 20(1), 13–23. DOI: <https://doi.org/10.1007/s00799-017-0228-4>
- Weiß, C., Schlecht, S. J., Rosenzweig, S., & Müller, M.** (2019). Towards measuring intonation quality of choir recordings: A case study on Bruckner's Locus Iste. In *Proceedings of the International Society for Music Information Retrieval Conference (ISMIR)*, pages 276–283. Delft, The Netherlands.
- Werner, N., Balke, S., Stöter, F.-R., Müller, M., & Edler, B.** (2017). trackswitch.js: A versatile webbased audio player for presenting scientific results. In *Proceedings of the Web Audio Conference (WAC)*, London, UK.
- Zalkow, F., Rosenzweig, S., Graulich, J., Dietz, L., Lemnaouar, E. M., & Müller, M.** (2018). A web-based interface for score following and track switching in choral music. In *Demos and Late Breaking News of the International Society for Music Information Retrieval Conference (ISMIR)*, Paris, France.
- Zapata, J. R., Davies, M. E. P., & Gómez, E.** (2014). Multi-feature beat tracking. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 22(4), 816–825. DOI: <https://doi.org/10.1109/TASLP.2014.2305252>

How to cite this article: Rosenzweig, S., Cuesta, H., Weiß, C., Scherbaum, F., Gómez, E., & Müller, M. (2020). Dagstuhl ChoirSet: A Multitrack Dataset for MIR Research on Choral Singing. *Transactions of the International Society for Music Information Retrieval*, 3(1), pp. 98–110. DOI: <https://doi.org/10.5334/tismir.48>

Submitted: 14 February 2020

Accepted: 10 June 2020

Published: 29 July 2020

Copyright: © 2020 The Author(s). This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International License (CC-BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. See <http://creativecommons.org/licenses/by/4.0/>.

 *Transactions of the International Society for Music Information Retrieval* is a peer-reviewed open access journal published by Ubiquity Press.

OPEN ACCESS 