Transactions of the International Society for
Music Information Retrieval

Qiuqiang Kong, Bochen Li, Jitong Chen, Yuxuan Wang (2020). GiantMIDI-Piano: A large-scale MIDI dataset for classical piano music, *Transactions of the International Society for Music Information Retrieval*, V(N), pp. xx–xx, DOI: https://doi.org/xx.xxxx/xxxx.xx

**ARTICLE TYPE**

# GiantMIDI-Piano: A large-scale MIDI dataset for classical piano music

Qiuqiang Kong*, Bochen Li*, Jitong Chen*, Yuxuan Wang*

## Abstract

Symbolic music datasets are important for music information retrieval and musical analysis. However, there is a lack of large-scale symbolic dataset for classical piano music. In this article, we create a GiantMIDI-Piano dataset containing 10,854 unique piano solo pieces composed by 2,786 composers. The dataset is collected as follows, we extract music piece names and composer names from the International Music Score Library Project (IMSLP). We search and download their corresponding audio recordings from the internet. We apply a convolutional neural network to detect piano solo pieces. Then, we transcribe those piano solo recordings to Musical Instrument Digital Interface (MIDI) files using our recently proposed high-resolution piano transcription system. Each transcribed MIDI file contains onset, offset, pitch and velocity attributes of piano notes, and onset and offset attributes of sustain pedals. GiantMIDI-Piano contains 34,504,873 transcribed notes, and contains metadata information of each music piece. To our knowledge, GiantMIDI-Piano is the largest classical piano MIDI dataset so far. We analyses the statistics of GiantMIDI-Piano including the nationalities, the number and duration of works of composers. We show the chroma, interval, trichord and tetrachord frequencies of six composers from different eras to show that GiantMIDI-Piano can be used for musical analysis. Our piano solo detection system achieves an accuracy of 89%, and the piano note transcription achieves an onset F1 of 96.72% evaluated on the MAESTRO dataset. GiantMIDI-Piano achieves an alignment error rate (ER) of 0.154 to the manually input MIDI files, comparing to MAESTRO with an alignment ER of 0.061 to the manually input MIDI files. We release the source code of acquiring the GiantMIDI-Piano dataset at `https://github.com/bytedance/GiantMIDI-Piano`.

**Keywords:** GiantMIDI-Piano, dataset, piano transcription

## 1. Introduction

Symbolic music datasets are important for music information retrieval (MIR) and musical analysis. In the past, musicians use musical notations to record music. Those notations include pitches, rhythms and chords of music. Musicologists used to analyses music pieces by reading those notations. Recently, computer has been used to process and analyses large-scale data, and has been widely used in MIR. However, there is a lack of large-scale symbolic music datasets covering a wide range of piano works in the past.

One difficulty of computer based MIR is that musical notations such as staffs are not directly readable by a computer. Therefore, converting music notations to computer readable formats are important. Early works of representing music in symbolic ways can be traced back to 1900s, where piano rolls (Bryner, 2002; Shi et al., 2019) were developed to record music that can be played on a musical instrument. Piano rolls are continuous rolls of paper with perforations punched into them. In 1981, Data Musical Instrument Digital Interface (MIDI) (Smith and Wood, 1981) was proposed as a technical standard to represent music, and is readable by computer. MIDI files use event messages to specify the instructions of music, including pitch, onset, offset and velocity of notes. MIDI files also carry rich information of music events such as sustain pedals. MIDI has been popular for music production in recent years.

In this work, we focus on building a large-scale MIDI dataset for classical piano music. Current piano MIDI datasets include the piano-midi.de (Krueger, 1996) dataset, the MAESTRO dataset (Hawthorne et al., 2019), the classical archives (Classical Archives,

*ByteDance

2000) dataset and the Kunstderfuge dataset (Kunstderfuge, 2002). However, even the largest dataset of those datasets is limited to hundreds of composers and hundreds hours of unique pieces (Kunstderfuge, 2002). On the other hand, MusicXML (Good et al., 2001) is another symbolic representation of music, but the scale of MusicXML datasets are smaller than current MIDI datasets. Optical music recognition (OMR) (Rebelo et al., 2012; Bainbridge and Bell, 2001) is a technique into transcribe image scores into symbolic representations. However, the performance of OMR systems are limited to image and score qualities.

In this article, we collect and transcribe a large-scale classical piano MIDI dataset called GiantMIDI-Piano dataset. The GiantMIDI-Piano dataset consists of 1,237 hours unique piano solo pieces composed by 2,786 composers. To our knowledge, GiantMIDI-Piano is the largest piano MIDI dataset so far. GiantMIDI-Piano is collected as follows: 1) We parse composer names and music piece names from the International Music Score Library Project (IMSLP)[1]; 2) For each music piece, we search its corresponding audio recording on YouTube; 3) We build a piano solo detection system to detect piano solo recordings; 4) Piano solo recordings are transcribed into MIDI files using our proposed high-resolution piano transcription system (Kong et al., 2020).

The GiantMIDI-Piano dataset can be used but not limited to the following research areas including: 1) Computer based musical analysis (Volk et al., 2011; Meredith, 2016); 2) Symbolic music generation (Yang et al., 2017; Hawthorne et al., 2019) and 3) Computer based music information retrieval (Casey et al., 2008; Choi et al., 2017) and expressive performance analysis (Cancino-Chacón et al., 2018). In this article, we analyses the statistics of GiantMIDI-Piano including the number of works, durations of works and nationalities of composers. In addition, we analyses the statistics of note, interval and chord distribution of six composers from different eras to show that GiantMIDI-Piano can be used for musical analysis. To evaluate the quality of GiantMIDI-Piano, we calculate the error rate (ER) of transcribed MIDI files by aligning them with manually input MIDI files.

This paper is organized as follows: Section 2 surveys piano MIDI datasets; Section 3 introduces the collection and transcription of GiantMIDI-Piano; Section 4 investigates the statistics of GiantMIDI-Piano; Section 5 evaluates the quality of GiantMIDI-Piano.

## 2. Dataset Survey

We introduce several piano MIDI datasets as follows. Piano-midi.de (Krueger, 1996) is a dataset of classical piano solo pieces. The author of piano-midi.de entered all notes using a MIDI sequencer. There are 571 pieces composed by 26 composers with a total duration of 36.7 hours of MIDI files in this dataset till Feb. 2020. The classical archives (Classical Archives, 2000) is a site containing a large number of MIDI files of classical music, including both piano and non piano pieces. There are 133 composers with a total duration of 46.3 hours of MIDI files in this dataset. Kunstderfuge (Kunstderfuge, 2002) is a large dataset containing piano solo and non piano solo works of 598 composers. All of the piano-midi.de, classical archive and Kunstderfuge datasets are entered using a MIDI sequencer, and are not real performances. The MAPS dataset (Emiya et al., 2010) used MIDI files from Piano-midi.de to render real recordings by playing back the MIDI files on a Yamaha Disklavier or using software synthesizers. The MAESTRO (Hawthorne et al., 2019) dataset contains over 200 hours of fine alignment MIDI files and audio recordings. In MAESTRO, virtuoso pianists performed on Yamaha Disklaviers with an integrated MIDI capture system. MAESTRO contains pieces from 62 composers. There are several duplicated pieces in MESTRO. For example, Scherzo No. 2 in B-flat Minor, Op. 31 composed by Chopin has been played 11 times. We manually remove the duplicated pieces when counting the number and duration of works. Table 1 shows the number of composers, the number of unique pieces, total durations and data types of different datasets. Data types include sequenced (Seq.) MIDI files, which are manually input MIDI files, and performance (Perf.) MIDI files from real performances. There are other MIDI datasets including the Lakh dataset (Raffel, 2016), the Bach Doodle dataset (Huang et al., 2019), the Bach Chorales dataset (Conklin and Witten, 1995), the URMP dataset (Li et al., 2018) and the Bach10 dataset (Duan et al., 2010). Google has collected 10,000 hours of piano recordings for music generation (Huang et al., 2018), while the dataset is not public available.

**Table 1:** Piano MIDI datasets

| Dataset | Composers | pieces | Hours | Type |
|---|---|---|---|---|
| piano-midi.de | 26 | 571 | 36.7 | Seq. |
| Classical archives | 133 | 856 | 46.3 | Seq. |
| Kunstderfuge | 598 | - | - | Seq. |
| MAESTRO | 62 | 529 | 84.3 | Perf. |
| MAPS | - | 270 | 18.6 | Perf. |
| **GiantMIDI-Piano** | **2,786** | **10,854** | **1,237** | **Live** |

## 3. GiantMIDI-Piano Dataset

### 3.1 Metadata from IMSLP

To begin with, we acquire the composer names and music piece names by parsing the webpages of the International Music Score Library Project (IMSLP, 2006), the largest public available music library in the world. In IMSLP, each composer has a webpage containing the list of his/her pieces. We acquire 143,701 music piece names composed by 18,067 composers by parsing those webpages. For each composer, if there ex-

[1] https://imslp.org

ists a biography link in the composer page, we access that biography link and search his/her birth, death and nationality of composers. We set the birth, death and nationality to "unknown" if a composer does not have such a biography link. We obtaine nationalities of 4,274 composers, and births of 5,981 composers out of 18,067 composers. We create a meta file containing the meta information of composers and music pieces. The attributes of metadata include surname, first name, piece name, birth, death and nationality.

### 3.2 Search Audio

We search audio recordings on YouTube by using keywords of "first name, surname, piece name" in the metadata. For each keyword, we select the first returned result on YouTube. There can be a case where the returned result may not exactly match the keyword. For example, for the keyword "Frédéric Chopin, Scherzo No.2 Op.31", the top returned result is "Chopin - Scherzo No. 2, Op. 31 (Rubinstein)". Although the keyword and the returned result are not perfectly matched, they indicate the same piece. We propose a Jaccard based similarity (Niwattanakul et al., 2013) to evaluate how a keyword and a returned result are matched. We denote the set of words in a keyword as $X$, and the set of words in a returned result as $Y$. The Jaccard similarity (Niwattanakul et al., 2013) between $X$ and $Y$ is defined as:

$$J = |X \cap Y| / |X|. \tag{1}$$

Higher $J$ indicates better match between $X$ and $Y$, and lower $J$ indicates less match between $X$ and $Y$. We set a similarity threshold to 0.6. If the similarity between $X$ and $Y$ is strictly larger than this threshold, then we say $X$ are $Y$ are matched, and vice versa. This similarity threshold is set empirically to balance the precision and recall of searched results. In total, we retrieve 60,724 audio recordings composed by 9,709 composers, out of 143,701 pieces composed by 18,067 composers.

### 3.3 Piano Solo Detection

The music pieces from IMSLP cover a wide range of instruments. We select only piano solo pieces to build the GiantMIDI-Piano dataset. Selecting pieces with keywords containing "piano" may lead to incorrect results. For example, a "Piano Concerto" is an ensemble of piano and orchestra, thus is not a piano solo. On the other hand, a keyword "Chopin, Frédéric, Nocturnes, Op.62" does not contain "piano", while it is indeed a piano solo. To address this problem, we train an audio based piano solo detection system using a convolutional neural network (CNN) (Kong et al., 2019). The piano detection system takes a 1-second segment as input. Log mel spectrograms are used as features, and are input to the CNN. The CNN consists of four convolutional layers. Each convolutional layer has a kernel size 3, and is followed by a batch normalization and a ReLU nonlinearity. The output of the CNN predicts the probability of the segment being a piano solo. Binary crossentropy is used as a loss function to optimize the CNN. We collect piano solo recordings as positive samples, and collect music and other sounds as negative samples. In addition, the mixtures of piano and other sounds are also used as negative samples. In inference, we average the predictions of all 1-second segments of an recording to obtain the piano solo probability. If the probability is strictly larger than 0.5, the audio recording is classified as a piano solo, and vice versa. In total, we obtain 10,854 piano solos out of 60,724 downloaded audio recordings, composed by 2,786 composers.

### 3.4 Piano Transcription

We transcribe all 10,854 piano solo recordings into MIDI files using our recently proposed high-resolution piano transcription system (Kong et al., 2020), an improvement to the onsets and frames piano transcription system (Hawthorne et al., 2017), (Hawthorne et al., 2019). The piano transcription system predicts all of pitch, onset, offset and velocity attributes of notes, and onset and offset attributes of sustain pedals. For piano note transcription, our system consists of a frame-wise classification, an onset regression, an offset regression and a velocity regression sub-module. Each sub-module is modeled with a convolutional recurrent neural network with eight convolutional layers and two bi-directional gated recurrent units (GRU) layers. The output of each module has a dimension of 88, equals to the number of notes on a piano. We model pedal onsets and offsets using the same sub-module architecture, except that there are only one output of the sub-module indicating onset or offset probabilities of pedals. In inference, all piano recordings are converted to monophonic with a sampling rate of 16 kHz. We use a short-time Fourier transform (STFT) with a Hanning window size 2048 and a hop size 160 to extract spectrograms, so there are 100 frames in a second. Then, mel filter banks with 229 bins are used to extract log mel spectrogram as input feature (Hawthorne et al., 2019). After input the log mel spectrogram to the trained piano transcription system, the system outputs the predicted probabilities of pitches, onsets, offsets and velocities. Finally, these outputs are post-processed to MIDI events. Our piano transcription system achieves a state-of-the-art onset F1 score of 96.72%, an onset & offset F1 score of 82.47% and an onset & offset & velocity F1 score of 80.92% on the test set of the MAESTRO dataset, and achieves a sustain-pedal onset F1 of 91.86%, and a sustain-pedal onset & offset F1 of 86.58%. Our piano transcription system outperforms previous onsets and frames Hawthorne et al. (2017, 2019) with on onset F1 score of 94.80%. The source code and pretrained models of our piano

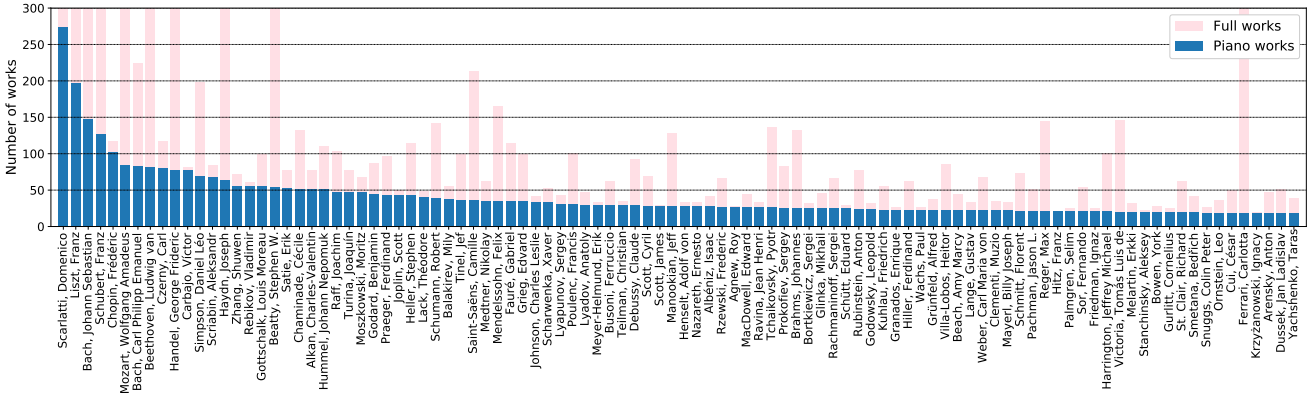


**Figure 1:** Number of works. Top 100 are shown.

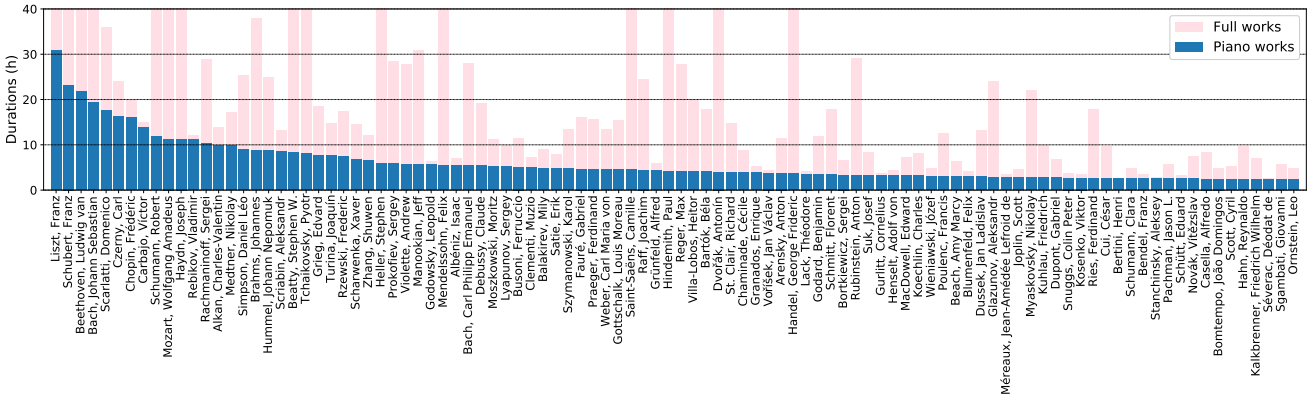


**Figure 2:** Duration of works. Top 100 are shown.

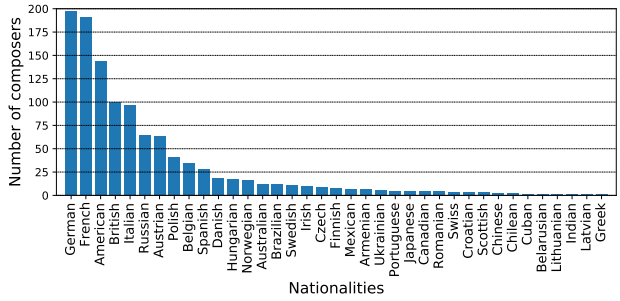


**Figure 3:** Number of composers with different nationalities.

transcription system (Kong et al., 2020) is released[2].

## 4. Statistics

We show the statistics of GiantMIDI-piano in this section. The statistics include the number and duration of pieces of different composers, the nationalities of composers, and the note distribution of composers. Then, we investigate the statistics of six composers from different eras by calculating their chroma, interval, trichord and tetrachord distributions.

[2] https://github.com/bytedance/piano_transcription

### 4.1 Number of Pieces

Fig. 1 shows the number of pieces composed by different composers sorted in descending order. The statistics of top 100 composers out of 2,786 composers are shown. Pink bars show the number of total pieces including both piano solo and non piano solo pieces. Blue bars show the number of piano solo pieces. Fig. 1 shows that there are 274 piano pieces composed by Scarlatti, followed 197 and 147 pieces composed by Liszt and J.S. Bach respectively. Some composers composed more piano solos than non piano solos, for example, Chopin composed 102 piano solos out of all 128 pieces. There are 226 composers composed more than 10 pieces in GiantMIDI-Piano. Fig. 1 shows that the numbers of pieces of different composers have a long tail distribution.

### 4.2 Duration of Pieces

Fig. 2 shows the duration of pieces of composer sorted in descending order. The durations differ from composers to composers. Liszt have the longest total piece durations of 30.88 hours, followed by Schubert of 23.22 hours and Beethoven of 21.82 hours. In GiantMIDI-Piano, there are 13 composers composed more than 10 hours, and 232 composers composed more than 1 hour. Handel composed 169.9 hours of music pieces, while only 3.7 hours of them are played on a modern piano. The ranks of composers in Fig. 2

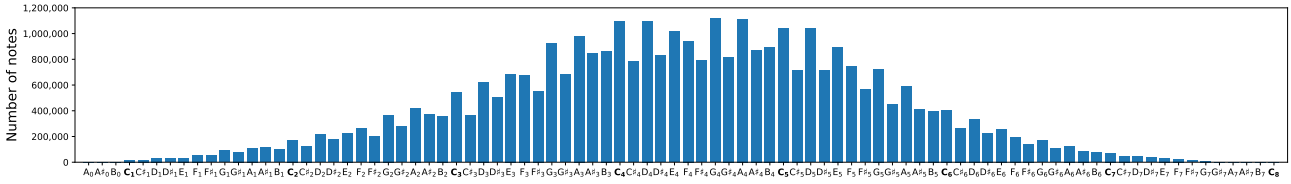


**Figure 4:** Note histogram of GiantMIDI-Piano.

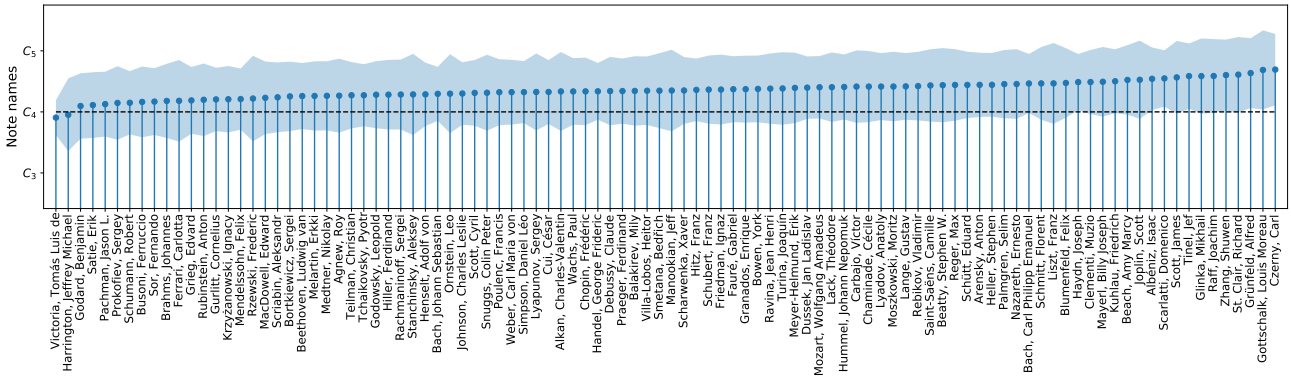


**Figure 5:** Note distribution of top 100 composers.

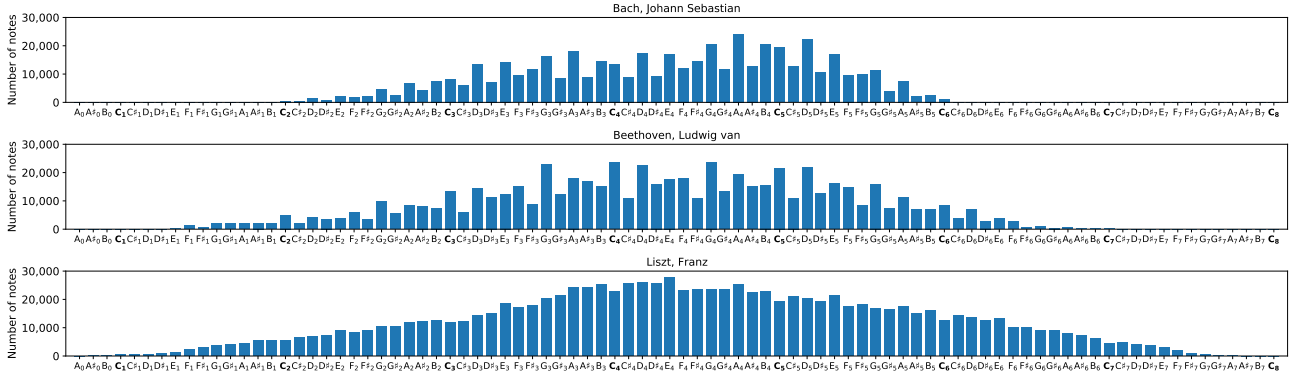


**Figure 6:** Notes histogram of J.S. Bach, Beethoven and Liszt.

are not the same as Fig. 1. For example, the average duration of Liszt is longer than Scarlatti.

### 4.3 Nationalities

Fig. 3 shows the number of composers sorted by nationalities in descending order. The nationalities of composers are obtained from Wikipedia. We ignore the composers that we could not find their nationalities, and label their nationalities as "unknown". We extract nationalities of 1,136 composers from 2,786 composers. Fig. 3 shows that there are 197 German composers, followed by 191 French composers and 144 American composers in GiantMIDI-Piano. The majority of composers acquired from ISMLP are from Europe. On the other hand, the number of composers from South America and Asia are fewer. For example, there are 12 composers from Brazil, 5 composers from Japan and 2 composers from China.

### 4.4 Notes Histogram

Fig. 4 shows the note histogram of GiantMIDI-Piano. In total, there are 34,504,873 transcribed notes. The horizontal axis shows the scientific pitch notations, which covers 88 notes on a modern piano from $A_0$ to $C_8$. Middle C is denoted as $C_4$. We do not distinguish enharmonic notes, for example, a note C♯/D♭ is simply denoted as C♯. Fig. 4 shows that the note histogram distributes according to a normal distribution, where the most played note is $G_4$. There are more notes near the $G_4$, and less notes far from $G_4$. The most played notes are within the octave between $C_4$ and $C_5$. White keys are being played more often than black keys. Fig. 5 shows the note distribution of top 100 composers from Fig. 1. Fig. 5 shows that the most played notes of composers are within $C_4$ and $C_5$. The shades show the one standard deviation area of note distributions. Tomás Luis de Victoria has the lowest average pitch of notes of $B_3$, whose works are mostly choir pieces played by a piano. Carl Czerny has the highest average

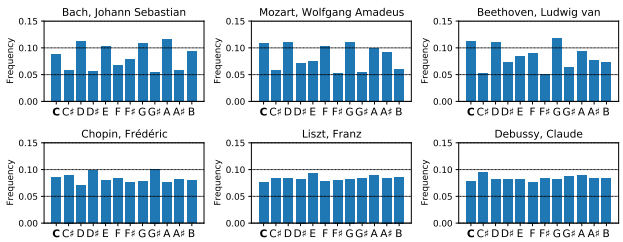


**Figure 7:** Chroma distribution of six composers.

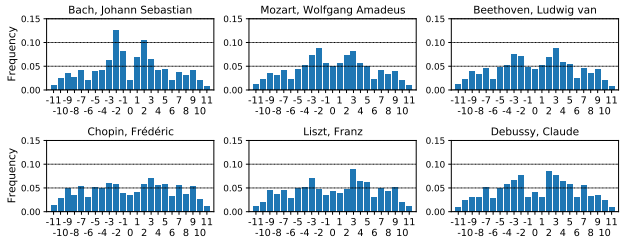


**Figure 8:** Interval distribution of six composers.

pitch of notes of G♯4. To further visualize the note distribution of different composers, Fig. 6 visualizes the note histogram of three composers from different eras, including J. S. Bach, Beethoven and Liszt. Fig. 6 shows that the note range of J. S. Bach are mostly between $F_2$ and $C_6$ covering four octaves, corresponding to the note range of a conventional harpsichord. Fig. 6 shows that the note range of Beethoven are mostly between $F_1$ and $C_7$, wider than the note range of J. S. Bach. On the other hand, Liszt has the widest note range, which covers the whole note range of a modern piano.

### 4.5 Chroma Distribution

Human perceive two musical notes are similar in color if they only differ by an octave. Twelve chroma values are represented by the set $\{C, C\sharp, D, D\sharp, E, F, F\sharp, G, G\sharp, A, A\sharp, B\}$. We use pitch-class sets (Forte, 1973) to represent chroma notes. That is, the notes from C to B are denoted as 0 to 11 respectively. We calculate the statistics of six composers from different eras, including J. S. Bach, Mozart, Beethoven, Chopin, Liszt and Debussy. Fig. 7 shows that J. S. Bach used $D, E, G, A$ most in his works. Mozart used $C, D, F, G$ most in their works, and used more A♯/B♭ than other composers. Beethoven preferred to use more $C, D, G$ than other notes. Chopin used D♯/E♭ and G♯/A♭ most in his works. Liszt and Debussy used all twelve notes more evenly in their works than other composers. Liszt used E most, and Debussy used C♯/D♭ most in their works.

### 4.6 Interval Distribution

An interval is the pitch difference between two notes. Intervals can be melodic intervals or harmonic intervals. We calculate intervals by parsing the MIDI files of GiantMIDI-Piano. An interval is calculated by $\Delta = y_n - y_{n-1}$, where $y_n$ is the midi pitch of a note, and $n$ is

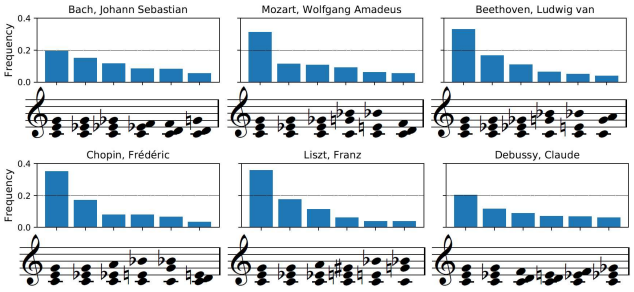


**Figure 9:** Trichord distribution of six composers.

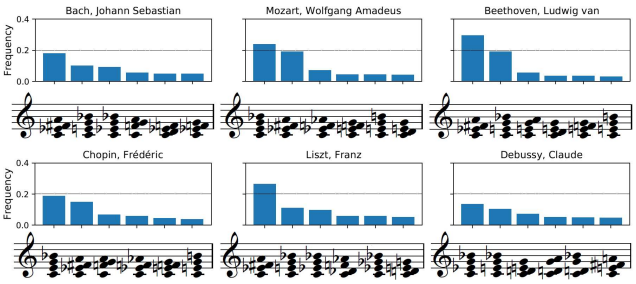


**Figure 10:** Tetrachord distribution of six composers.

the index of the note. We calculate ordered pitch-class intervals (Forte, 1973), that is, positive intervals and negative intervals are different. For example, the interval $\Delta$ for an upward progress from $C_4$ to $D_4$ is 2, and the interval $\Delta$ for a downward progress from $C_4$ to $A_3$ is $-3$. We only calculate the intervals within an octave, from -11 to 11. For example, the value 11 indicates a major seventh interval. Fig. 8 shows the interval distribution of the six composers. J. S. and Debussy used major second most in their works. Mozart, Beethoven, Chopin and Liszt used most minor third in their works. The interval distribution are not symmetric, for example, J. S. Bach used more downward major second than upward major second. Major seventh and tritone are least used by all composers.

### 4.7 Trichord Distribution

We adopt the set musical theory (Forte, 1973) to analyses the chord distribution in GiantMIDI-Piano. A trichord is a set of any three pitch-classes (Forte, 1973). We only calculate the trichords with onsets within 50 ms. Using the set musical theory, a major triad can be written as $\{0, 4, 7\}$, where the interval between 0 and 4 is a major third, and the interval between 4 and 7 is a minor third. Following (Forte, 1973), two chords are equivalent if and only if they are reducible to the same prime form by transposition and inversion. So we convert all chords to their prime form, for example, a chord $\{2, 6, 9\}$ will be reduced to $\{0, 4, 7\}$. Fig. 9 shows the trichord distribution of the six composers. All composers used major triad $\{0, 4, 7\}$ most in their pieces. J. S. Bach used more minor triad $\{0, 3, 7\}$ than other composers. Liszt used used more augmented triad $\{0, 4, 8\}$ than other composers. Debussy used $\{0, 2, 5\}$, $\{0, 2, 4\}$ and $\{0, 3, 5\}$ more, which distinguished him from other

composers.

### 4.8 Tetrachord Distribution

A tetrachord is a set of any four pitch-classes (Forte, 1973). Similar to trichords, we only calculate the tetrachords with onsets within 50 ms. For example, dominant seventh chord can be denoted as $\{0,4,7,10\}$. Fig. 10 shows the tetrachord distribution of the six composers. Fig. 10 shows that Bach, Beethoven and Liszt used diminished seventh $\{0,3,6,9\}$ most, while Mozart and Chopin used dominant seventh most $\{0,4,7,10\}$. Beethoven used more diminished seventh $\{0,3,6,9\}$ and dominant seventh $\{0,4,7,10\}$ than other composers. The second inversion of half-diminished seventh $\{0,3,7,9\}$ is also a common used tetrachord of Bach, Liszt and Debussy. The tetrachord distribution of Debussy is different from other composers, where different tetrachords are more evenly distributed. Minor seventh $\{0,3,7,10\}$ is mostly used by Debussy. Other tetrachords that distinguish Debussy from other composers include $\{0,2,4,7\}$ and $\{0,2,7,9\}$.

## 5. Evaluate Quality of GiantMIDI-Piano

### 5.1 Piano Solo Evaluation

To evaluate the precision of piano solo detection, we manually label 100 piano solos out of 10,854 music recordings in GiantMIDI-Piano. There are 89 piano solo recordings out of 100 recordings are indeed piano solos, indicating a precision of 89% of the piano solo detection system. Out of the 11 false positives, there are eight guitar solos that are classified as piano solos. Precision is more important factor than recall when building GiantMIDI-Piano.

### 5.2 GiantMIDI-Piano Evaluation

Evaluating the piano transcription quality of GiantMIDI-Piano is a challenging problem because there are no ground truth MIDI files of real recordings for evaluation. To address this problem, we propose an alignment method to evaluate the quality of transcribed MIDI files. We call manually input MIDI files sequenced MIDI files. For a same music piece, we align the sequenced MIDI file with the performance MIDI file by using a hidden Markov model (HMM) based alignment tool (Nakamura et al., 2017). Usually, the sequenced MIDI file and performance MIDI file are not perfectly aligned caused by system alignment error and performance error. The total error consists of insertion (I), deletion (D) and substitution (S) error. We use error rate (ER) (Mesaros et al., 2016) to evaluate the total error between a sequenced MIDI file and a performance MIDI file:

$$ER = \frac{S + D + I}{N}, \tag{2}$$

where $N$ is the number of reference notes. For MIDI files from the MAESTRO dataset Hawthorne et al. (2019), the total error $e_1$ comes from system error and performance error:

$$e_1 = e_{\text{system}} + e_{\text{performance}}, \tag{3}$$

where the performance error $e_{\text{performance}}$ comes from that a pianist may miss or add pitches accidentally while performing (Repp, 1996). The system error $e_{\text{system}}$ is caused by the alignment tool (Nakamura et al., 2017). For MIDI files from the transcribed GiantMIDI-Piano dataset, there is an extra transcription error $e_{\text{transcription}}$ introduced by the piano transcription system. The total error is:

$$e_2 = e_{\text{performance}} + e_{\text{transcription}} + e_{\text{system}}. \tag{4}$$

For a music piece, we first search its performance MIDI in the test set of MAESTRO dataset to calculate $e_1$. Then, we search its corresponding transcribed MIDI files in GiantMIDI-Piano to calculate $e_2$. Then, we compare $e_2$ with $e_1$, and propose a *relative error* as:

$$r = e_2 - e_1. \tag{5}$$

The relative error is used as an approximation to the transcription ER of GiantMIDI-Piano. This approximation is based on the assumption that the values of $e_{\text{performance}}$ and $e_{\text{system}}$ in $e_1$ and $e_2$ are similar. Lower $r$ suggests better transcription quality. We evaluate 52 music pieces that appear in both the test set of MAESTRO and GiantMIDI-Piano, and search their sequenced MIDI files on the internet. Long music pieces such as Sonatas are split into movements before applying the alignment algorithm. Table 2 shows the alignment performance. MAESTRO achieve median alignment S, D, I and ER of 0.009, 0.024, 0.021 and 0.061 respectively. GiantMIDI-Piano achieve median alignment S, D, I and ER of 0.015, 0.051, 0.069 and 0.154 respectively. The relative error $r$ between MAESTRO and GiantMIDI-Piano is 0.094. The first column of Fig. 11 shows the box plot metrics of MAESTRO. The outliers in the figure are mostly come from different interpretations of trills and tremolos. The second column of Fig. 11 shows the box plot metrics of GiantMIDI-Piano. In GiantMIDI-Piano, Keyboard Sonata in E-Flat Major, Hob. XVI/49 composed Haydn achieves the lowest ER of 0.037, while Prelude and Fugue in A-flat major, BWV 862 composed by Bach achieves the highest ER of 0.679 (outlier beyond the plot range). This underperformance is caused by the piece is not played by

**Table 2:** Alignment performance of performance MIDI and sequenced MIDI.

| | D | I | S | ER |
|---|---|---|---|---|
| Maestro ground truth | 0.009 | 0.024 | 0.018 | 0.061 |
| GiantMIDI-Piano | 0.015 | 0.051 | 0.069 | 0.154 |
| Relative difference | 0.006 | 0.026 | 0.047 | 0.094 |

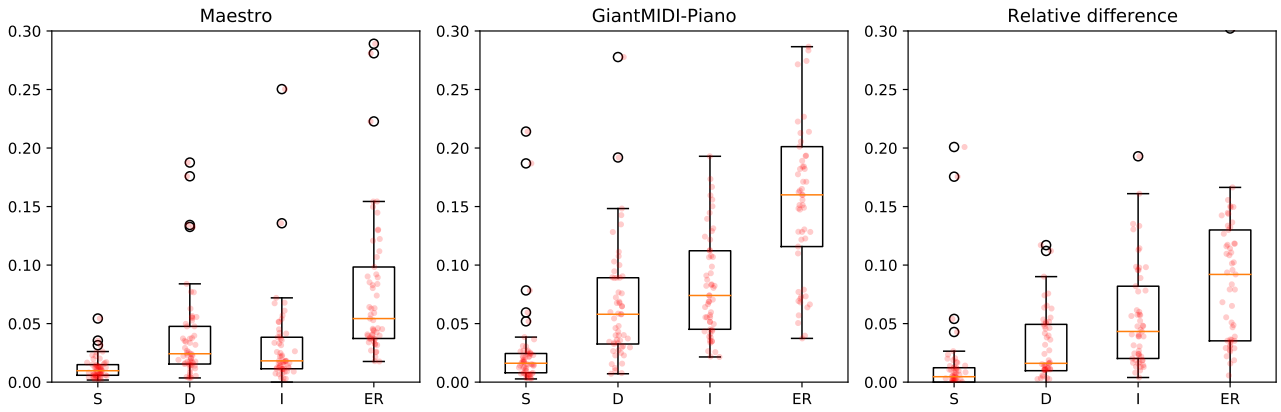


**Figure 11:** From left to right: ER of 52 pieces in the MAESTRO dataset; ER of 52 pieces in the GiantMIDI-Piano dataset; Relative ER between the MAESTRO and GiantMIDI-Piano dataset.

standard pitch with $A_4$ 440 Hz. The third column of Fig. 11 shows the relative ER between MAESTRO and GiantMIDI-Piano. The relative median scores of S, D, I and ER are 0.006, 0.026, 0.047 and 0.094 respectively, suggesting there are less deletions than insertions.

## 6. Conclusion

We collect and transcribe a large-scale GiantMIDI-Piano dataset containing 10,854 unique classical piano pieces composed by 2,786 composers. The total duration of GiantMIDI-Piano is 1,237 hours, and there are 34,504,873 transcribed piano notes. GiantMIDI-Piano are transcribed from audio recordings searched from YouTube using meta information extracted from IMSLP. To our knowledge, GiantMIDI-Piano is the largest piano MIDI dataset so far. The piano solo detection system used in GiantMIDI-Piano achieves an accuracy of 89%, and the piano transcription system achieves a relative error rate of 0.094. We have released the source code for acquiring GiantMIDI-Piano. In future, GiantMIDI-Piano can be used but not limited to musical analysis, music generation and music information retrieval.

## References

Bainbridge, D. and Bell, T. (2001). The challenge of optical music recognition. *Computers and the Humanities*, 35(2):95–121.

Bryner, B. (2002). *The piano roll: a valuable recording medium of the twentieth century*. PhD thesis, Department of Music, University of Utah.

Cancino-Chacón, C. E., Grachten, M., Goebl, W., and Widmer, G. (2018). Computational models of expressive music performance: A comprehensive and critical review. *Frontiers in Digital Humanities*, 5:25.

Casey, M. A., Veltkamp, R., Goto, M., Leman, M., Rhodes, C., and Slaney, M. (2008). Content-based music information retrieval: Current directions and future challenges. *Proceedings of the IEEE*, 96(4):668–696.

Choi, K., Fazekas, G., Cho, K., and Sandler, M. (2017). A tutorial on deep learning for music information retrieval. *arXiv preprint arXiv:1709.04396*.

Classical Archives (2000). Classical Archives. `www.classicalarchives.com`.

Conklin, D. and Witten, I. H. (1995). Multiple viewpoint systems for music prediction. *Journal of New Music Research*, 24(1):51–73.

Duan, Z., Pardo, B., and Zhang, C. (2010). Multiple fundamental frequency estimation by modeling spectral peaks and non-peak regions. *IEEE Transactions on Audio, Speech, and Language Processing*, 18(8):2121–2133.

Emiya, V., Bertin, N., David, B., and Badeau, R. (2010). Maps-a piano database for multipitch estimation and automatic transcription of music. In *Research Report, INRIA-00544155f*.

Forte, A. (1973). *The structure of atonal music*, volume 304. Yale University Press.

Good, M. et al. (2001). Musicxml: An internet-friendly format for sheet music. In *XML conference and expo*, pages 03–04. Citeseer.

Hawthorne, C., Elsen, E., Song, J., Roberts, A., Simon, I., Raffel, C., Engel, J., Oore, S., and Eck, D. (2017). Onsets and frames: Dual-objective piano transcription. *arXiv preprint arXiv:1710.11153*.

Hawthorne, C., Stasyuk, A., Roberts, A., Simon, I., Huang, C. A., Dieleman, S., Elsen, E., Engel, J., and Eck, D. (2019). Enabling factorized piano music modeling and generation with the maestro dataset. *International Conference on Learning Representations (ICLR)*.

Huang, C.-Z. A., Hawthorne, C., Roberts, A., Dinculescu, M., Wexler, J., Hong, L., and Howcroft, J. (2019). The Bach Doodle: Approachable music composition with machine learning at scale. In *International Society for Music Information Retrieval (ISMIR)*.

Huang, C.-Z. A., Vaswani, A., Uszkoreit, J., Simon, I., Hawthorne, C., Shazeer, N., Dai, A. M., Hoffman, M. D., Dinculescu, M., and Eck, D. (2018). Music transformer: Generating music with long-term structure. In *International Conference on Learning Representations*.

IMSLP (2006). International Music Score Library Project. `imslp.org`.

Kong, Q., Cao, Y., Iqbal, T., Xu, Y., Wang, W., and Plumbley, M. D. (2019). Cross-task learning for audio tagging, sound event detection and spatial localization: DCASE 2019 baseline systems. In *arXiv:1904.03476*.

Kong, Q., Li, B., Song, X., Wan, Y., and Wang, Y. (2020). High-resolution piano transcription with pedals by regressing onsets and offsets times. In *arXiv preprint arXiv:2010.01815*.

Krueger, B. (1996). Classical Piano MIDI Page. `http://www.piano-midi.de`.

Kunstderfuge (2002). Kunstderfuge. `http://www.kunstderfuge.com`.

Li, B., Liu, X., Dinesh, K., Duan, Z., and Sharma, G. (2018). Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications. *IEEE Transactions on Multimedia*, 21(2):522–535.

Meredith, D. (2016). *Computational music analysis*, volume 62. Springer.

Mesaros, A., Heittola, T., and Virtanen, T. (2016). Metrics for polyphonic sound event detection. *Applied Sciences*, 6(6):162.

Nakamura, E., Yoshii, K., and Katayose, H. (2017). Performance error detection and post-processing for fast and accurate symbolic music alignment. In *International Society for Music Information Retrieval (ISMIR)*, pages 347–353.

Niwattanakul, S., Singthongchai, J., Naenudorn, E., and Wanapu, S. (2013). Using of jaccard coefficient for keywords similarity. In *Proceedings of the International MultiConference of Engineers and Computer Scientists (IMECS)*, pages 380–384.

Raffel, C. (2016). *Learning-based methods for comparing sequences, with applications to audio-to-midi alignment and matching*. PhD thesis, Columbia University.

Rebelo, A., Fujinaga, I., Paszkiewicz, F., Marcal, A. R., Guedes, C., and Cardoso, J. S. (2012). Optical music recognition: state-of-the-art and open issues. *International Journal of Multimedia Information Retrieval*, 1(3):173–190.

Repp, B. H. (1996). The art of inaccuracy: Why pianists' errors are difficult to hear. *Music Perception*, 14(2):161–183.

Shi, Z., Sapp, C. S., Arul, K., McBride, J., and Smith III, J. O. (2019). Supra: Digitizing the stanford university piano roll archive. In *International Society for Music Information Retrieval (ISMIR)*, pages 517–523.

Smith, D. and Wood, C. (1981). The'usi', or universal synthesizer interface. In *Audio Engineering Society Convention 70*.

Volk, A., Wiering, F., and K., P. K. (2011). Unfolding the potential of computational musicology. In *International Conference on Informatics and Semiotics in Organisations (ICISO)*, pages 137–144.

Yang, L., Chou, S., and Yang, Y. (2017). MidiNet: A convolutional generative adversarial network for symbolic-domain music generation. In *International Society for Music Information Retrieval (ISMIR)*, pages 324–331.