

Supplementary Materials for M4Singer

A Datasheet

In this section we copy the questions from *Datasheets for Datasets*¹ and answer them.

A.1 Motivation

The questions in this section are primarily intended to encourage dataset creators to clearly articulate their reasons for creating the dataset and to promote transparency about funding interests. The latter may be particularly relevant for datasets created for research purposes.

- **Q: For what purpose was the dataset created?** Was there a specific task in mind? Was there a specific gap that needed to be filled? Please provide a description.

A: we present a free-to-use Multi-style, Multi-singer Mandarin singing collection with elaborately annotated Musical scores as well as its benchmarks. It aims to provide a publicly available corpus for multi-singer singing voice synthesis (SVS) related researches. Thus different SVS approaches which use this corpus could be easy to reproduce, and make fair comparisons between each other.

- **Q: Who created the dataset (e.g., which team, research group) and on behalf of which entity (e.g., company, institution, organization)?**

A: Zhejiang University.

- **Q: Who funded the creation of the dataset?** If there is an associated grant, please provide the name of the grantor and the grant name and number.

A: The singing voice collection is funded by Zhejiang University, while the cost on label annotation is by Huawei Noah's Ark Lab.

- **Q: Any other comments?**

A: No.

A.2 Composition

Dataset creators should read through these questions prior to any data collection and then provide answers once data collection is complete. Most of the questions in this section are intended to provide dataset consumers with the information they need to make informed decisions about using the dataset for their chosen tasks. Some of the questions are designed to elicit information about compliance with the EU's General Data Protection Regulation (GDPR) or comparable regulations in other jurisdictions.

- **Q: What do the instances that comprise the dataset represent (e.g., documents, photos, people, countries)?** Are there multiple types of instances (e.g., movies, users, and ratings; people and interactions between them; nodes and edges)? Please provide a description.

A: Singing voice, alignment and musical score.

- **Q: How many instances are there in total (of each type, if appropriate)?**

A: About 30 hours of high-quality Mandarin singing voice with annotation.

- **Q: Does the dataset contain all possible instances or is it a sample (not necessarily random) of instances from a larger set?** If the dataset is a sample, then what is the larger set? Is the sample representative of the larger set (e.g., geographic coverage)? If so, please describe how this representativeness was validated/verified. If it is not representative of the larger set, please describe why not (e.g., to cover a more diverse range of instances, because instances were withheld or unavailable).

A: It contains all possible instances.

¹<https://arxiv.org/abs/1803.09010>

- **Q: What data does each instance consist of?** “Raw” data (e.g., unprocessed text or images) or features? In either case, please provide a description.

A: Raw data, but it is segmented at sentence-level to facilitate training.
- **Q: Is there a label or target associated with each instance?** If so, please provide a description.

A: Singing voice with alignment and musical score annotation.
- **Q: Is any information missing from individual instances?** If so, please provide a description, explaining why this information is missing (e.g., because it was unavailable). This does not include intentionally removed information, but might include, e.g., redacted text.

A: No.
- **Q: Are relationships between individual instances made explicit (e.g., users’ movie ratings, social network links)?** If so, please describe how these relationships are made explicit.

A: No.
- **Q: Are there recommended data splits (e.g., training, development/validation, testing)?** If so, please provide a description of these splits, explaining the rationale behind them.

A: Yes, we preset several tasks by preparing training set and test set.
- **Q: Are there any errors, sources of noise, or redundancies in the dataset?** If so, please provide a description.

A: The singing voices are recorded in professional studio, and they are clean enough for high-fidelity singing voice synthesis. The labeling process is carried out by professionals with musical backgrounds. If there are any errors, they would exist in the labels of duration and pitch, due to the slight difference in subjective perception.
- **Q: Is the dataset self-contained, or does it link to or otherwise rely on external resources (e.g., websites, tweets, other datasets)?** If it links to or relies on external resources, a) are there guarantees that they will exist, and remain constant, over time; b) are there official archival versions of the complete dataset (i.e., including the external resources as they existed at the time the dataset was created); c) are there any restrictions (e.g., licenses, fees) associated with any of the external resources that might apply to a dataset consumer? Please provide descriptions of all external resources and any restrictions associated with them, as well as links or other access points, as appropriate.

A: Self-contained.
- **Q: Does the dataset contain data that might be considered confidential (e.g., data that is protected by legal privilege or by doctor– patient confidentiality, data that includes the content of individuals’ non-public communications)?** If so, please provide a description.

A: No.
- **Q: Does the dataset contain data that, if viewed directly, might be offensive, insulting, threatening, or might otherwise cause anxiety?** If so, please describe why.

A: No.
- **Q: Does the dataset identify any subpopulations (e.g., by age, gender)?** If so, please describe how these subpopulations are identified and provide a description of their respective distributions within the dataset.

A: No.
- **Q: Is it possible to identify individuals (i.e., one or more natural persons), either directly or indirectly (i.e., in combination with other data) from the dataset?** If so, please describe how.

A: No.

- **Q: Does the dataset contain data that might be considered sensitive in any way (e.g., data that reveals race or ethnic origins, sexual orientations, religious beliefs, political opinions or union memberships, or locations; financial or health data; biometric or genetic data; forms of government identification, such as social security numbers; criminal history)?** If so, please provide a description.

A: No.

- **Q: Any other comments?**

A: No.

A.3 Collection Process

As with the questions in the previous section, dataset creators should read through these questions prior to any data collection to flag potential issues and then provide answers once collection is complete. In addition to the goals outlined in the previous section, the questions in this section are designed to elicit information that may help researchers and practitioners to create alternative datasets with similar characteristics. Again, questions that apply only to datasets that relate to people are grouped together at the end of the section.

- **Q: How was the data associated with each instance acquired?** Was the data directly observable (e.g., raw text, movie ratings), reported by subjects (e.g., survey responses), or indirectly inferred/derived from other data (e.g., part-of-speech tags, model-based guesses for age or language)? If the data was reported by subjects or indirectly inferred/derived from other data, was the data validated/verified? If so, please describe how.

A: Singing voice is recorded in a professional recording studio.

- **Q: What mechanisms or procedures were used to collect the data (e.g., hardware apparatuses or sensors, manual human curation, software programs, software APIs)?** How were these mechanisms or procedures validated?

A: Recording studio.

- **Q: If the dataset is a sample from a larger set, what was the sampling strategy (e.g., deterministic, probabilistic with specific sampling probabilities)?**

A: N/A.

- **Q: Who was involved in the data collection process (e.g., students, crowdworkers, contractors) and how were they compensated (e.g., how much were crowdworkers paid)?**

A: Professional singers.

- **Q: Over what timeframe was the data collected?** Does this timeframe match the creation timeframe of the data associated with the instances (e.g., recent crawl of old news articles)? If not, please describe the timeframe in which the data associated with the instances was created.

A: N/A.

- **Q: Were any ethical review processes conducted (e.g., by an institutional review board)?** If so, please provide a description of these review processes, including the outcomes, as well as a link or other access point to any supporting documentation.

A: Yes.

- **Q: Did you collect the data from the individuals in question directly, or obtain it via third parties or other sources (e.g., websites)?**

A: From the individuals.

- **Q: Were the individuals in question notified about the data collection?** If so, please describe (or show with screenshots or other information) how notice was provided, and provide a link or other access point to, or otherwise reproduce, the exact language of the notification itself.

A: Yes. All singers sign an agreement and agree to open source their singing voice for academic usage prior to collection.

- **Q: Did the individuals in question consent to the collection and use of their data?** If so, please describe (or show with screenshots or other information) how consent was requested and provided, and provide a link or other access point to, or otherwise reproduce, the exact language to which the individuals consented.

A: Yes.

- **Q: If consent was obtained, were the consenting individuals provided with a mechanism to revoke their consent in the future or for certain uses?** If so, please provide a description, as well as a link or other access point to the mechanism (if appropriate).

A: Yes.

- **Q: Has an analysis of the potential impact of the dataset and its use on data subjects (e.g., a data protection impact analysis) been conducted?** If so, please provide a description of this analysis, including the outcomes, as well as a link or other access point to any supporting documentation.

A: No.

- **Q: Any other comments?**

A: No.

A.4 Preprocessing/cleaning/labeling

Dataset creators should read through these questions prior to any preprocessing, cleaning, or labeling and then provide answers once these tasks are complete. The questions in this section are intended to provide dataset consumers with the information they need to determine whether the “raw” data has been processed in ways that are compatible with their chosen tasks. For example, text that has been converted into a “bag-of-words” is not suitable for tasks involving word order.

- **Q: Was any preprocessing/cleaning/labeling of the data done (e.g., discretization or bucketing, tokenization, part-of-speech tagging, SIFT feature extraction, removal of instances, processing of missing values)?** If so, please provide a description. If not, you may skip the remaining questions in this section. documentation.

A: We performed data cleaning and segmentation work to facilitate usage.

- **Q: Was the “raw” data saved in addition to the preprocessed/cleaned/labeled data (e.g., to support unanticipated future uses)?** If so, please provide a link or other access point to the “raw” data.

A: Yes.

- **Q: Is the software that was used to preprocess/clean/label the data available?** If so, please provide a link or other access point.

A: Yes. We will release the preprocessing scripts that are used to build the benchmarks with this dataset.

- **Q: Any other comments?**

A: No.

A.5 Uses

The questions in this section are intended to encourage dataset creators to reflect on the tasks for which the dataset should and should not be used. By explicitly highlighting these tasks, dataset creators can help dataset consumers to make informed decisions, thereby avoiding potential risks or harms.

- **Q: Has the dataset been used for any tasks already?**

A: Yes, it has been presented in the paper.

- **Q: Is there a repository that links to any or all papers or systems that use the dataset?**

A: This has been presented in the paper.

- **Q: What (other) tasks could the dataset be used for?**

A: singing voice synthesis related and music information retrieval related works.

- **Q: Is there anything about the composition of the dataset or the way it was collected and preprocessed/cleaned/labeled that might impact future uses?** For example, is there anything that a dataset consumer might need to know to avoid uses that could result in unfair treatment of individuals or groups (e.g., stereotyping, quality of service issues) or other risks or harms (e.g., legal risks, financial harms)? If so, please provide a description. Is there anything a dataset consumer could do to mitigate these risks or harms?

A: No.

- **Q: Are there tasks for which the dataset should not be used?** If so, please provide a description.

A: No.

- **Q: Any other comments?**

A: No.

A.6 Distribution

Dataset creators should provide answers to these questions prior to distributing the dataset either internally within the entity on behalf of which the dataset was created or externally to third parties

- **Q: Will the dataset be distributed to third parties outside of the entity (e.g., company, institution, organization) on behalf of which the dataset was created?** If so, please provide a description.

A: No.

- **Q: How will the dataset will be distributed (e.g., tarball on website, API, GitHub)?** Does the dataset have a digital object identifier (DOI)?

A: We adopt the application form for data management.

- **Q: When will the dataset be distributed?**

A: We have provided data access methods.

- **Q: Will the dataset be distributed under a copyright or other intellectual property (IP) license, and/or under applicable terms of use (ToU)?** If so, please describe this license and/or ToU, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms or ToU, as well as any fees associated with these restrictions.

A: CC BY-NC-SA 4.0.

- **Q: Have any third parties imposed IP-based or other restrictions on the data associated with the instances?** If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any relevant licensing terms, as well as any fees associated with these restrictions.

A: No.

- **Q: Do any export controls or other regulatory restrictions apply to the dataset or to individual instances?** If so, please describe these restrictions, and provide a link or other access point to, or otherwise reproduce, any supporting documentation.

A: No.

- **Q: Any other comments?**

A: No.

A.7 Maintenance

As with the questions in the previous section, dataset creators should provide answers to these questions prior to distributing the dataset. The questions in this section are intended to encourage dataset creators to plan for dataset maintenance and communicate this plan to dataset consumers.

- **Q: Who will be supporting/hosting/maintaining the dataset?**
A: Zhejiang University.
- **Q: How can the owner/curator/manager of the dataset be contacted (e.g., email address)?**
A: apply4m4singer@gmail.com
- **Q: Is there an erratum?** If so, please provide a link or other access point.
A: apply4m4singer@gmail.com
- **Q: Will the dataset be updated (e.g., to correct labeling errors, add new instances, delete instances)?** If so, please describe how often, by whom, and how updates will be communicated to dataset consumers (e.g., mailing list, GitHub)?
A: Yes, all the updates will be synced on the website.
- **Q: If the dataset relates to people, are there applicable limits on the retention of the data associated with the instances (e.g., were the individuals in question told that their data would be retained for a fixed period of time and then deleted)?** If so, please describe these limits and explain how they will be enforced.
A: No.
- **Q: Will older versions of the dataset continue to be supported/hosted/maintained?** If so, please describe how. If not, please describe how its obsolescence will be communicated to dataset consumers.
A: No, all the updates will be synced on the website.
- **Q: If others want to extend/augment/build on/contribute to the dataset, is there a mechanism for them to do so?** If so, please provide a description. Will these contributions be validated/verified? If so, please describe how. If not, why not? Is there a process for communicating/distributing these contributions to dataset consumers? If so, please provide a description.
A: CC BY-NC-SA 4.0.
- **Q: Any other comments?**
A: No.