



# SingStyle111: A Multilingual Singing Dataset With Style Transfer

Contact: [shuqid@cs.cmu.edu](mailto:shuqid@cs.cmu.edu)  
<https://shuqid.net>

Shuqi Dai, Siqu Chen, Yuxuan Wu, Ruxin Diao, Roy Huang, Roger B. Dannenberg

## Key Points

- A large, studio-quality, monophonic singing dataset
- 111 songs(224 versions), 8 singers, 12.8 hours, 3 languages
- Various singing styles (including creative improvisations)
- **Style transfer** demonstration: 80 songs covering at least two distinct singing styles performed by the same singer
- Detailed **human annotations**: phrase segmentation, lyrics phoneme-audio alignment, performance MIDI, score
- Applications: Singing Voice Synthesis (SVS), Singing Voice Conversion (SVC), Singing Style Transfer, Singing Transcription, Score Following, Expressive Performance, etc.
- Publicly free for research: [shuqid.net/singingstyle111](https://shuqid.net/singingstyle111)

## Motivation

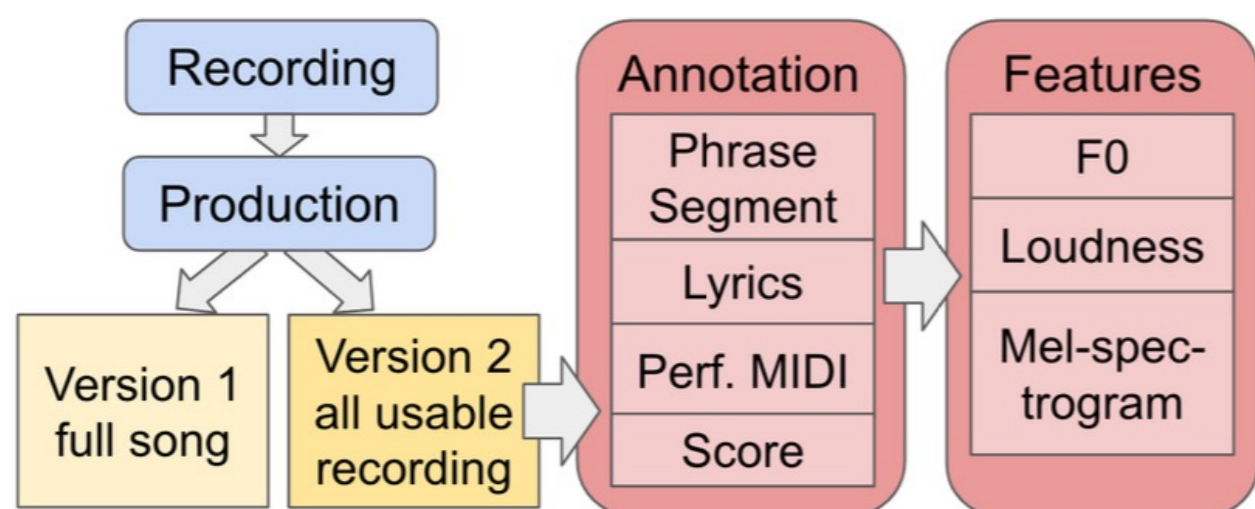
We lack **high-quality, large-scale, diverse, annotated, public** data for singing voice research.

Dataset	Language	Style	#Hour	#Singer	Quality	Musicality	Score	Align-ment	Style Transfer
Opencpop [41]	Chinese	Pop	5.25	1	Studio	Ama.	Perform. MIDI	✓	✗
M4Singer [42]	Chinese	Pop	29.77	20	Studio	50% Ama. 50% Prof.	Perform. MIDI	✓	✗
Children Song [43]	Korean English	Children	4.86	1	Studio	Prof. but plain	Perform. MIDI	word	✗
Tohoku Kiritan [44]	Japanese	Pop	0.95	1	Studio	Prof.	Score	✓	✗
PopCS [28]	Chinese	Pop	5.89	6	Not Clean	Ama.	✗	✗	✗
Open-Singer [35]	Chinese	Pop	50	66	Studio	Ama.	✗	✗	✗
VocalSet [45]	Five Vowels	Opera	10.1	20	Studio	Prof.	Score	✓	technique transfer
NHSS [47]	English	Pop	3.5	10	Studio	Ama.	✗	✓	✗
NUS-48E [48]	English	Pop Children	1.41	12	Studio	Ama.	✗	✓	✗
RWC [49]	Japanese English	Pop	4	27	Not Solo	Prof.	Both	✗	✗
TONAS [50]	Spanish	Flamenco	0.34	> 40	Not Clean	Prof.	✗	✗	✗
Vocadito [51]	Seven Languages	Pop Children	0.23	29	Not Clean	Ama.	✗	✗	✗
MIR-1K [52]	Chinese	Pop	2.22	19	Not Solo	Ama.	✗	✗	✗
<b>StyleSing111 (Ours)</b>	English Chinese Italian	Opera Pop Folk Jazz etc.	12.8	8	Studio	Prof.	Both	✓	✓

A comparison of existing singing datasets. Score means if there is score or performance MIDI file provided. "Perform. MIDI" = "Performance MIDI". "Both" means both performance MIDI files synchronized with the singing audio and sheet music scores are provided. Alignment means whether or not there is duration annotation at the phoneme level for lyrics and audio. "Ama." = "Amateur," "Prof." = "Professional."

## Dataset Overview

### Data Collection Pipeline

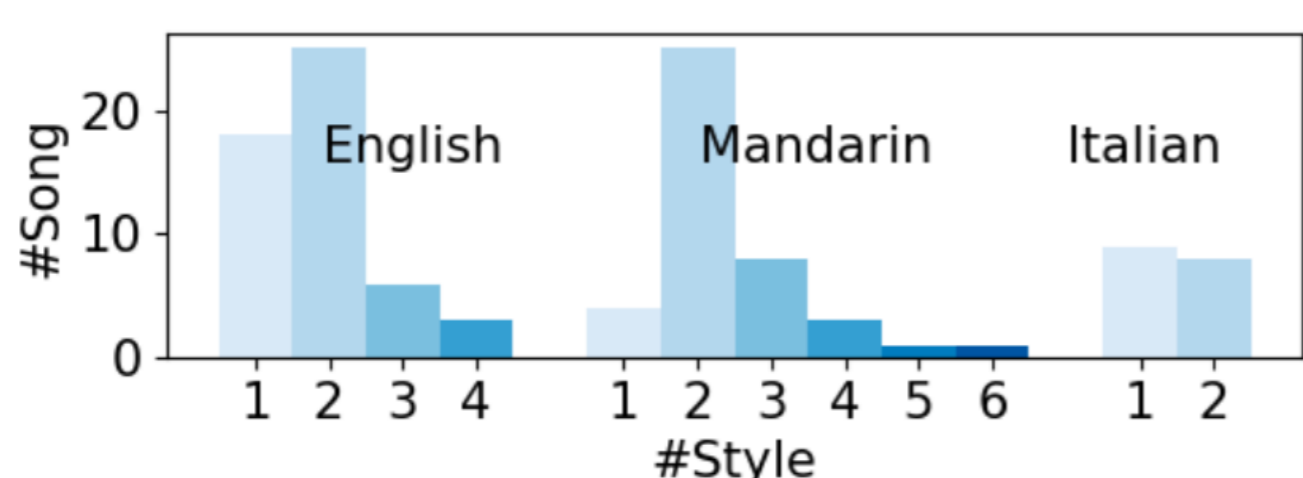


Version 1: edited full-length songs

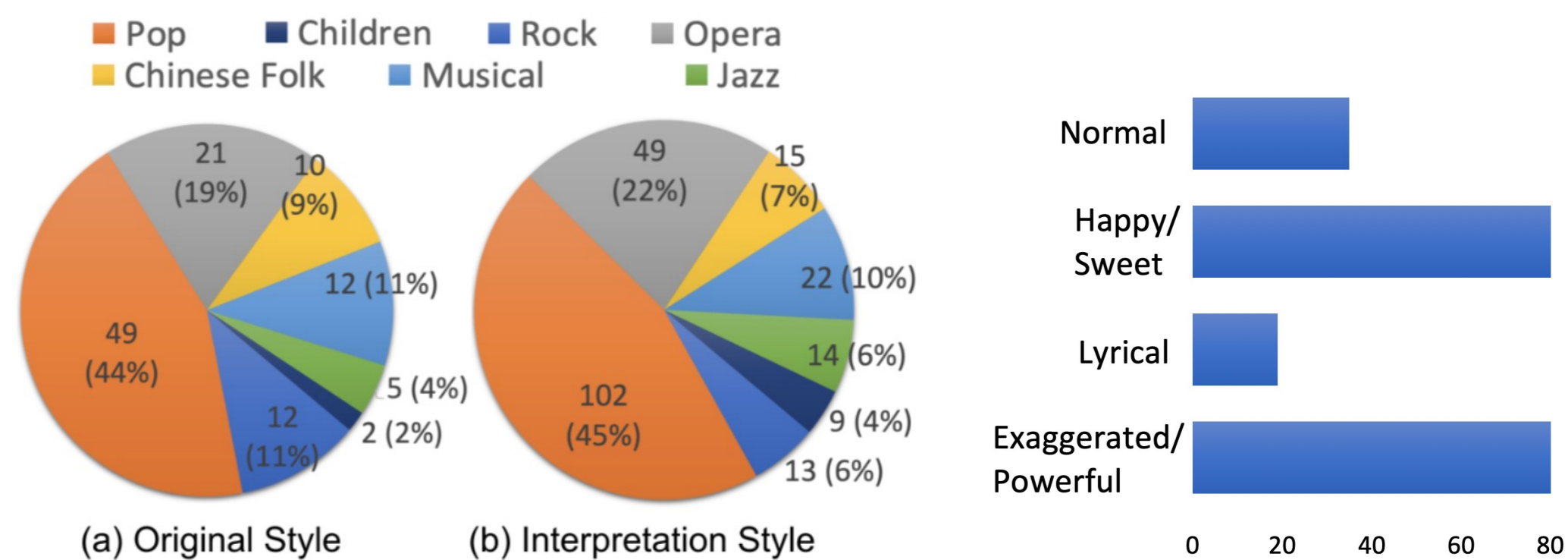
Version 2: all usable recordings of the song in segments, including redos

- **Repertoire:** English-372 min. Chinese-307 min. Italian-88 min.

224 style demonstrations. E.g., English songs have 18 songs with one style version, 25 songs with 2 different styles, 6 songs with 3 styles, and 3 songs with 4 styles.



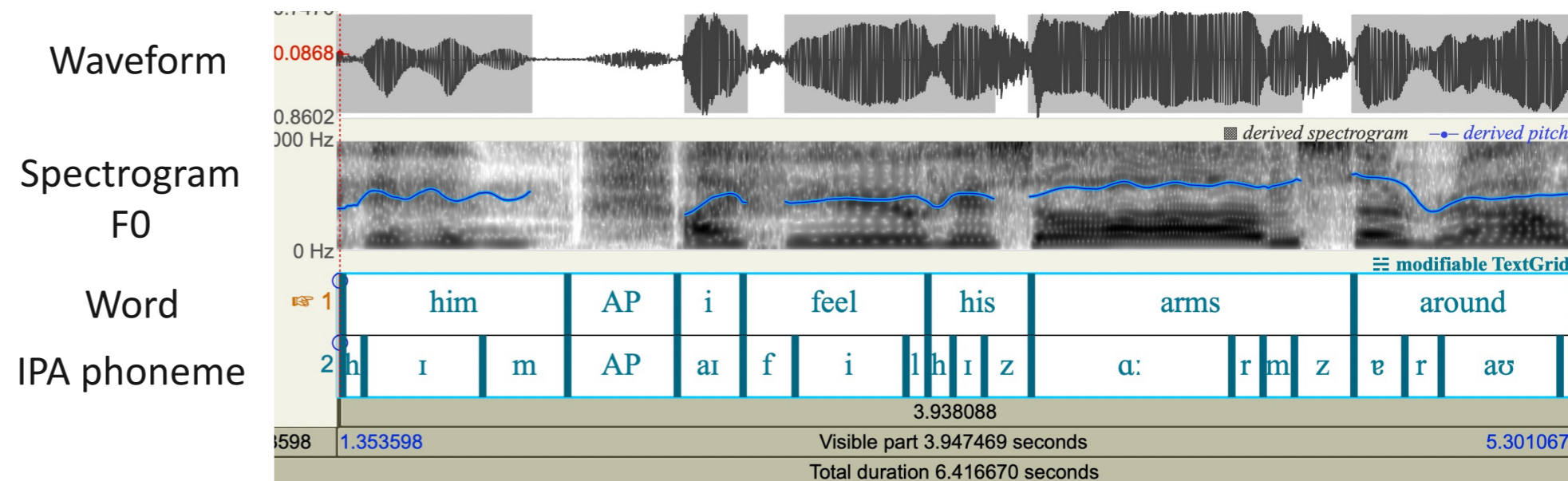
- **Style Distribution:** 7 style genres + 4 emotion modes



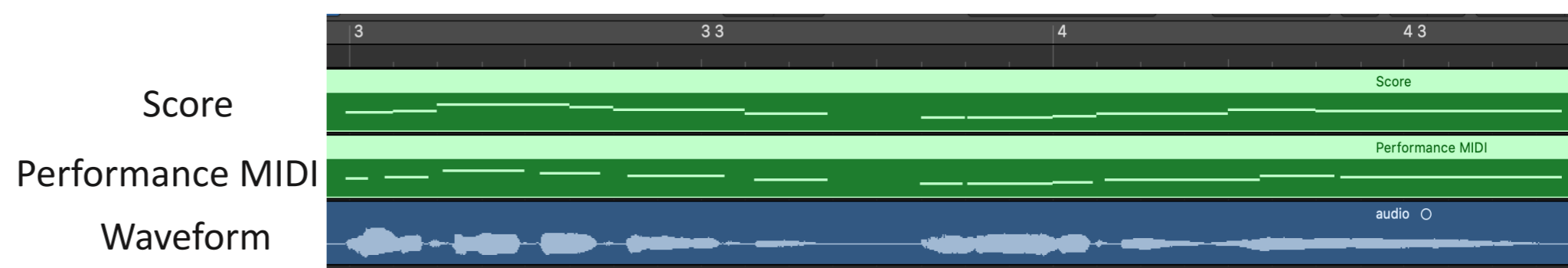
## Data Annotation & Features

**Multiple rounds of human annotation:** we started a long-term music data annotation project, hired and trained people to label music data. Need music data annotation? contact us!

- Example of phoneme-audio alignment in a phrase



- Example of performance MIDI and score in a phrase



- Extracted features: Mel-spectrogram, F0, Loudness curves

## New Work: Expressive Singing Voice Synthesis

Work @ NVIDIA Research Internship

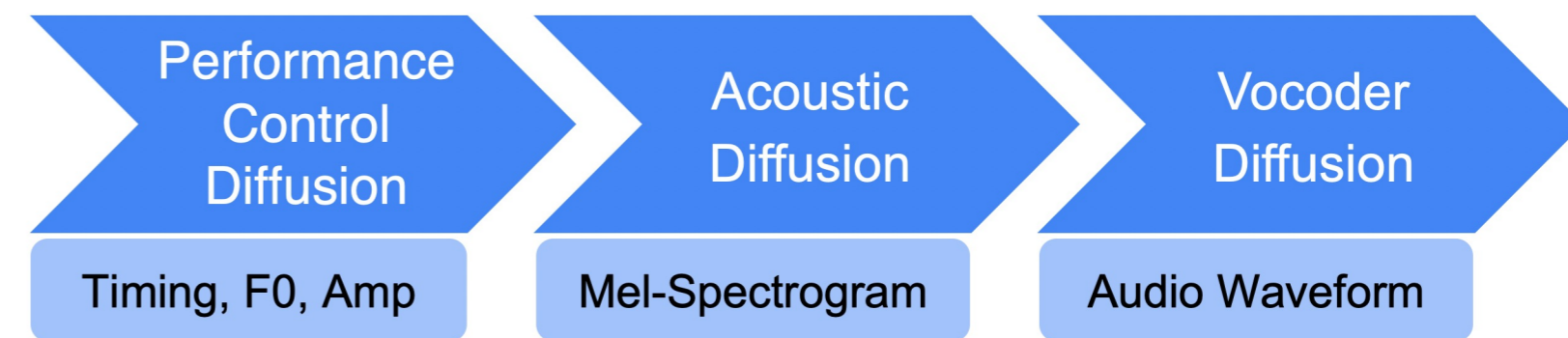
Input: 1. Score; 2. Lyrics; 3. Style; 4. Singer ID

Output: Expressive and realistic singing

Method: 1. Five **Diffusion** Models

2. Focus on **Expressive Performance Control**
- "No GOOD Control, No GOOD Synthesis"

Demo



## New Work: Zero-shot Singing Synthesis

Work @ Adobe Research Internship

Input: 1. Score & Lyrics & Style

2. 5-second speech audio of target (unseen target voice in training data)

Output: Singing in target's voice

Demo



## Shuqi is on the job market

I am completing my PhD by summer 2024 and looking for opportunities in both academia and industry. If you're aware of any openings, I'd appreciate a heads-up! Thank you!

CV

