Curating an A Cappella Dataset for Source Separation

Ting-Yu Pan*, Kexin Phyllis Ju*, Hao-Wen Dong

Background

- A cappella music presents unique challenges for source separation due to its diverse vocal styles and the presence of vocal percussion.
- Current a cappella datasets are limited in size and diversity (20 - 40 songs), hindering the development of robust source separation models.

A Cappella Dataset

**55 Studio-Quality
A cappella Songs**
Mandarin, English, Korean, and Hakka Chinese

Golden Dataset

AI Augmentation
Pitch Shifting, Voice Cloning, Voice Synthesis
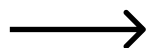
Augmented Dataset

MIDI and MusicXML

Symbolic Dataset

Two-Step Source Separation Pipeline

Vocal Percussion (VP) Extraction
We configured Demucs to output four stems and treat its "drums" channel as VP.

→

Vocal Harmony Separation
We will apply UMSS to decompose the "VP-less" residual into SATB parts.

Preliminary Experiment

Fine-tune Source Separation baseline model - "Demucs" for VP extraction
Our fine-tuned model improved VP SDR from 5.22 dB to 7.62 dB **(a 2.4 dB gain)**, outperforming both the drum-finetuned and official baselines without harming harmonic separation.

Model	VP	Other	All
Pretrained (official)	5.22	10.66	7.94
Pretrained (drum)	3.66	9.24	6.45
Fine-tuned (ours)	**7.62**	**11.63**	**9.62**

Why A Cappella Source Separation?

Group Recording → Source Separation →

Soprano
Alto
Tenor
Bass
VP
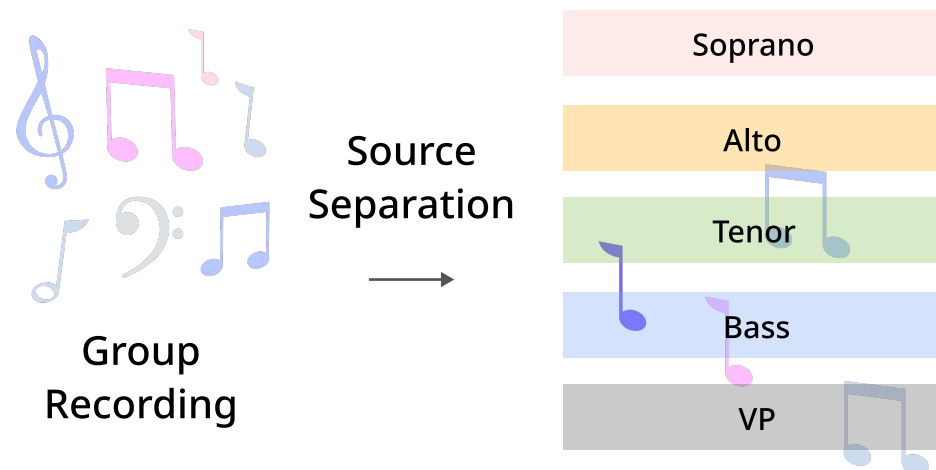
Example Use Case:
A Cappella Training System