# A Study of Audio Mixing Methods for Piano Transcription in Violin-Piano Ensembles

Hyemi Kim[b,h]   Jiyun Park[#]   Taegyun Kwon[#]   Dasaem Jeong[x]   Juhan Nam[b,#]

[b] Kim Jaechul Graduate School of AI, KAIST, Daejeon, South Korea | [h] Electronics and Telecommunications Research Institute, Daejeon, South Korea

| [#] Graduate School of Culture Technology, KAIST, Daejeon, South Korea | [x] Sogang University, Seoul, South Korea

Piano transcription in Violin-Piano Ensembles
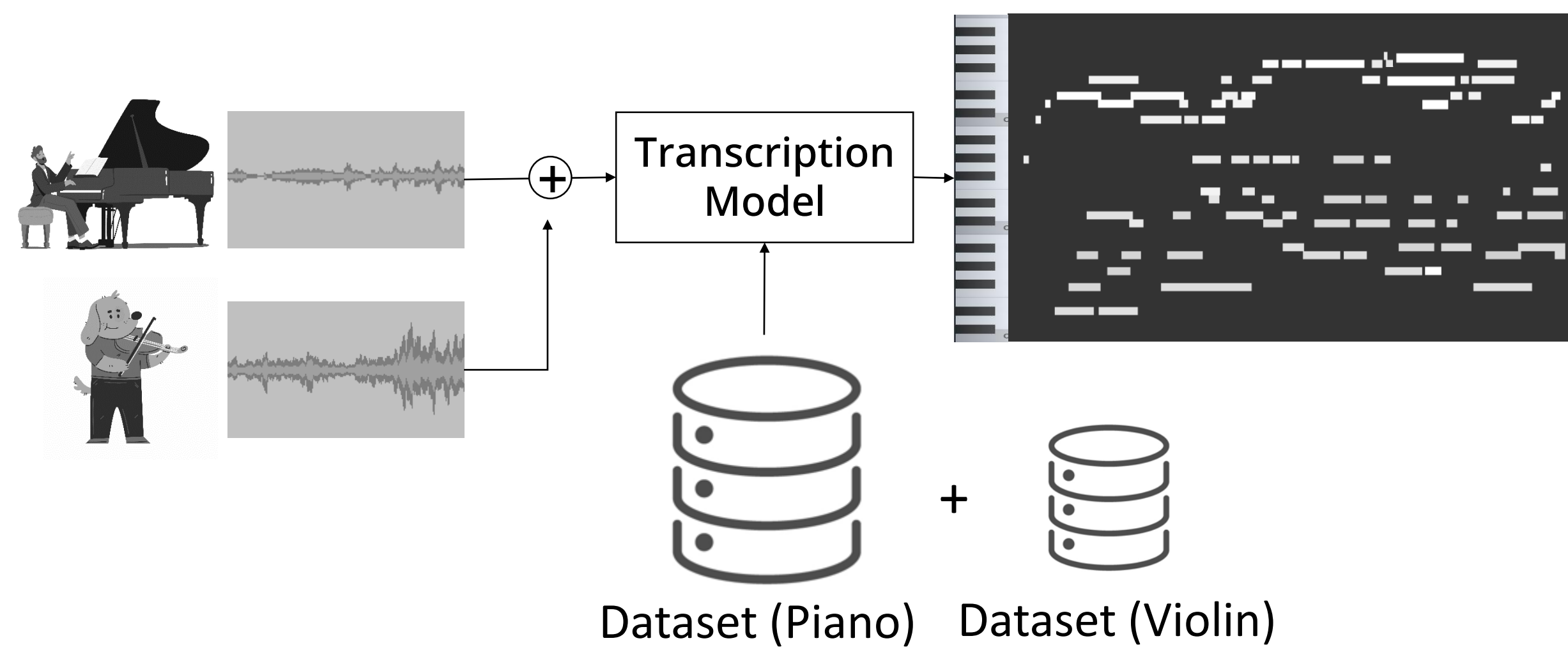
PFVN-synth Dataset

- Triplet audios (piano, violin, mixture)
- Midi labels (piano, violin)
- 45 pieces by 23 classical composers
- 7 hours
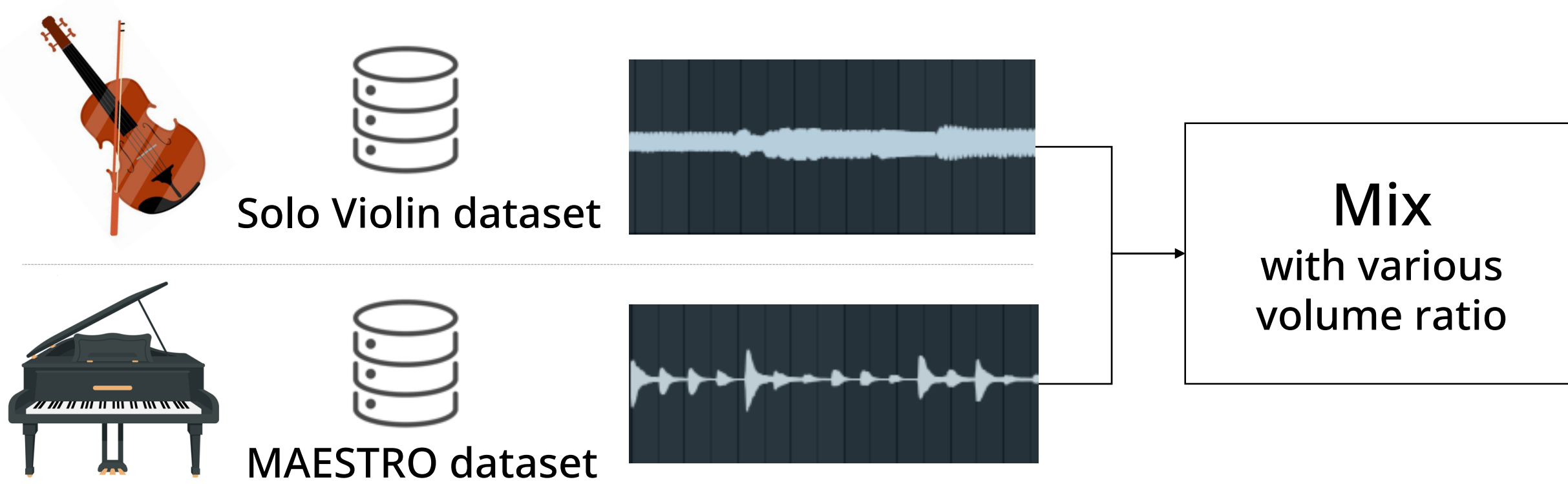- Mono standard CD quality
  - 44.1kHz, 16-bit
- 3 splits

Split	Pieces	Duration(h)
Train	32	5.7
Validation	3	0.5
Test	10	0.8

https://zenodo.org/record/7703620

Real Performance Dataset

- Solo violin pieces
  - 48 audio recordings from YouTube composed by 7 composers
  - 4.6 hours
  - train set : validation set = 9:1
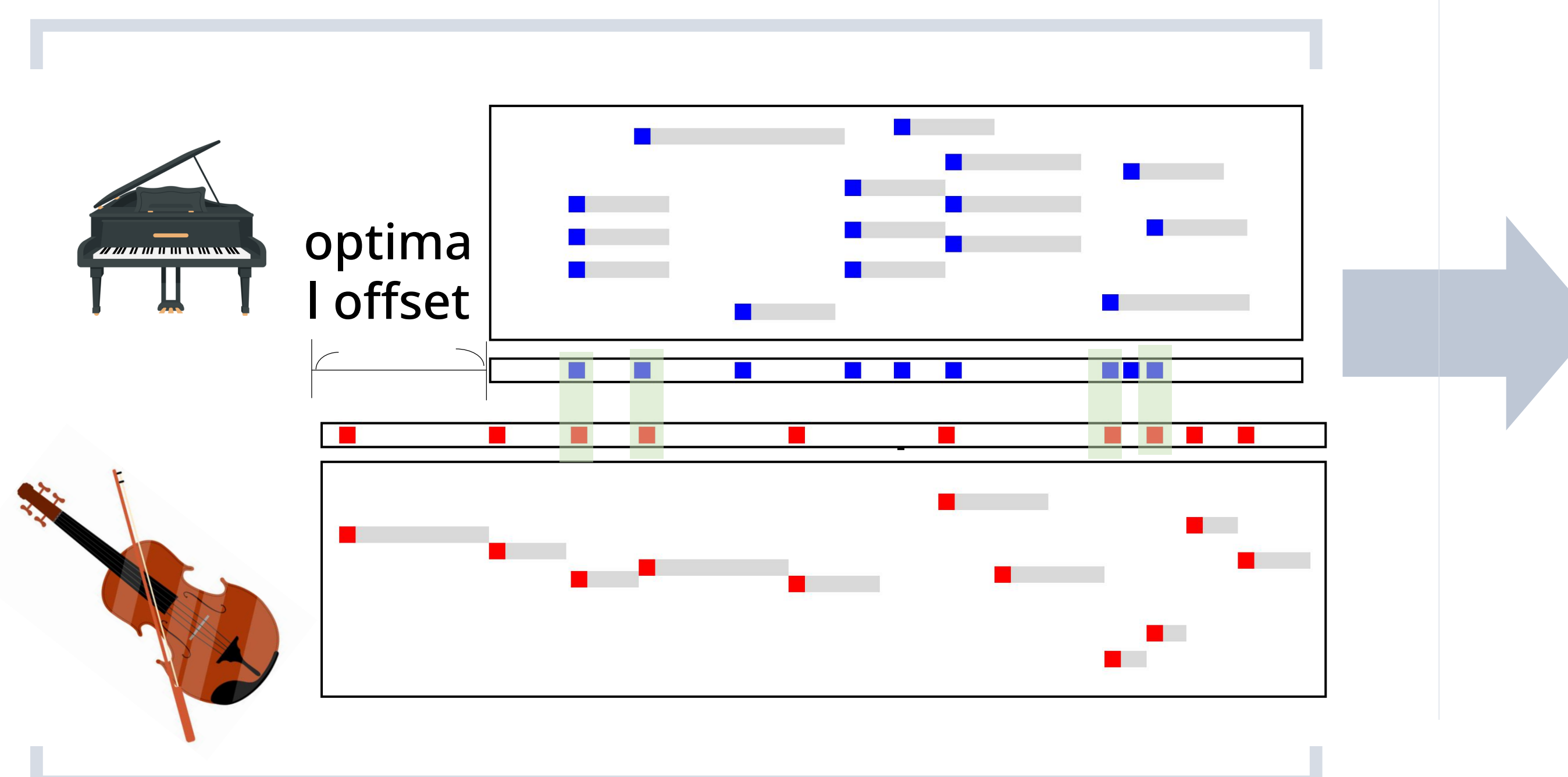

Proposed Method

Key-based Mixing (Harmonic characteristics)

- Key estimation: `madmom.features.key`
- Harmonically-related keys
  - Same / Dominant / Subdominant key
  - Major mode : E – E/A/B/Am
  - Minor mode : Am – Am/E/Dm/Em
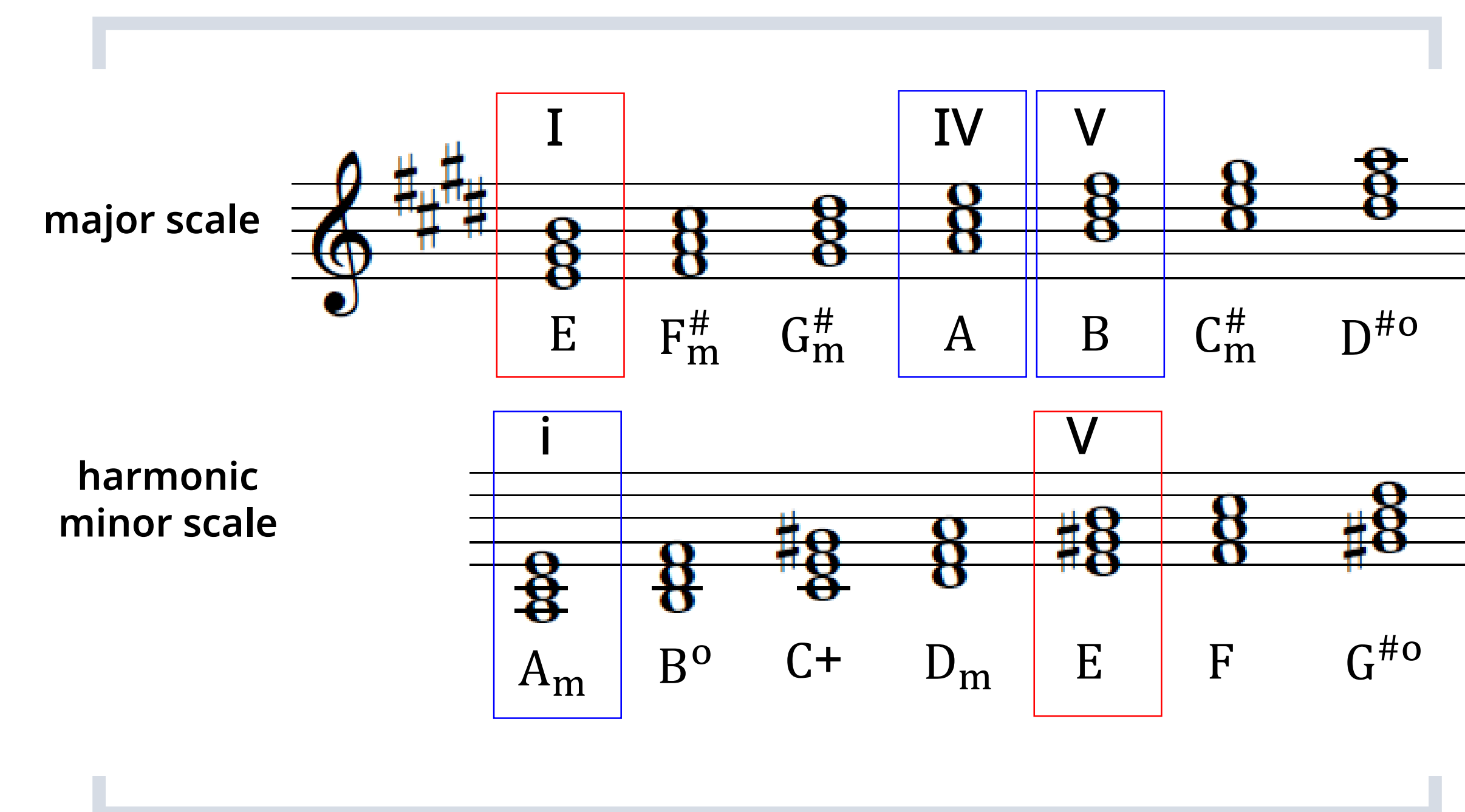
Onset-based Mixing (Temporal characteristics)

- Mix audio excerpts that aligned note onsets as closely as possible
  - We shifted the entire piano clip by adjusting the global offset.
- Onset label
  - MIDI label (for synthesized dataset)
  - Onset estimation using `librosa.onset.onset_detect` (for real recordings)

Experimental Results

Experiment Settings

- Network structure: *Onsets and Frames*
- Metric: *mir_eval*

Mixing Methods

- Random mixing (Baseline)
- Original pair (only for PFVN-synth)
- Proposed methods
  - Onset-based Mixing
  - Key-based Mixing
  - Key and Onset-based Mixing
- Lower bound
  - evaluate violin-piano ensembles using the model trained with piano only
- Upper bound
  - evaluate piano audio alone using the model trained with piano only

Experiments – PFVN-synth Dataset

Mixing method	Frame			Note with onset			Note with offset			Note w. offset & vel		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
Lower bound	82.0	51.7	62.7	78.3	62.2	68.5	51.9	41.0	45.3	40.8	32.6	35.7
Original pair	90.8	81.8	85.7	99.1	96.8	97.9	72.6	71.0	71.8	71.4	69.8	70.6
Random mixing	95.4	86.6	90.6	99.2	95.5	97.2	80.2	77.2	78.6	78.9	75.9	77.3
Onset-based mixing	96.0	**88.9**	**92.1**	99.3	97.1	**98.2**	82.3	80.5	81.3	80.8	**79.0**	**79.9**
Key-based mixing	**96.2**	88.1	91.8	**99.4**	**97.2**	**98.2**	81.7	79.9	80.7	80.2	78.5	79.3
Key and onset-based mixing	95.5	87.9	91.4	99.3	96.1	97.6	80.4	77.8	79.1	79.2	76.7	77.9
Upper bound	97.5	90.9	94.0	99.2	97.8	98.5	86.3	85.1	85.7	85.3	84.2	84.7

Experiments – Real Performance Dataset

Mixing method	Frame			Note with onset			Note with offset			Note w. offset & vel		
	P	R	F1	P	R	F1	P	R	F1	P	R	F1
PFVN-synth												
Lower bound	71.8	62.5	66.2	63.5	75.0	68.4	36.4	43.1	39.2	26.3	31.1	28.3
Random mixing	82.3	**75.4**	78.4	**92.5**	81.5	86.4	57.6	50.7	53.9	45.6	40.2	42.6
Onset-based mixing	83.4	73.9	78.1	89.8	81.1	85.0	57.6	52.0	54.5	45.1	40.9	42.7
Key-based mixing	**85.2**	75.1	**79.6**	92.3	**83.1**	**87.3**	**59.0**	**53.0**	**55.7**	**46.5**	**41.9**	**44.0**
Key and Onset-based mixing	82.9	74.5	78.2	90.2	81.0	85.1	57.0	51.1	53.8	44.6	40.2	42.2
Upper bound	78.7	87.7	82.7	89.3	93.0	91.0	56.1	58.6	57.3	43.9	45.9	44.8
MusicNetEM												
Lower bound	75.1	50.3	59.7	71.8	80.8	75.9	28.9	32.2	30.4	-	-	-
Random mixing	78.0	**58.2**	**65.8**	87.7	78.4	82.7	36.7	32.9	34.7	-	-	-
Onset-based mixing	78.9	57.4	65.6	87.4	78.4	82.4	**37.1**	33.4	35.0	-	-	-
Key-based mixing	**79.2**	56.8	65.5	**87.7**	**79.4**	**83.2**	**37.1**	33.7	35.3	-	-	-
Key and Onset-based mixing	**79.2**	56.6	65.1	87.1	79.1	82.8	37.0	**33.8**	35.3	-	-	-

Conclusion – Musically-informed mixing methods are more effective than random mixing.