INTERNATIONAL AUDIO LABORATORIES ERLANGEN
A joint institution of Fraunhofer IIS and Universität Erlangen-Nürnberg

A Cross-Version Approach to Audio Representation Learning for Orchestral Music

Michael Krause[1], Christof Weiß[2], Meinard Müller[1]

[1]International Audio Laboratories Erlangen, Germany, [2]University of Würzburg, Germany

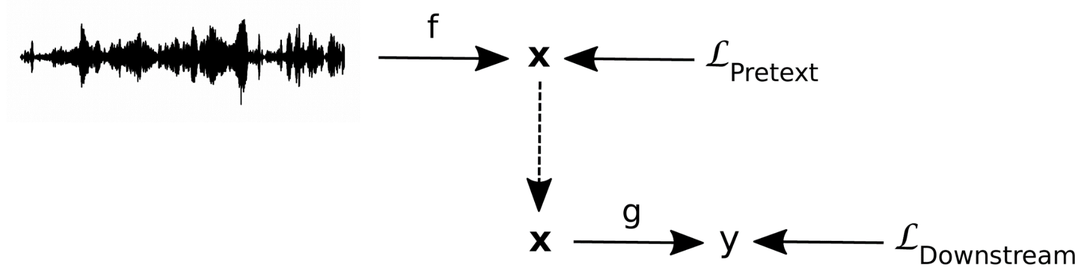
ISMIR 2023 Milan, Italy
Nov. 5-9, 2023
POLITECNICO MILANO 1863

Summary

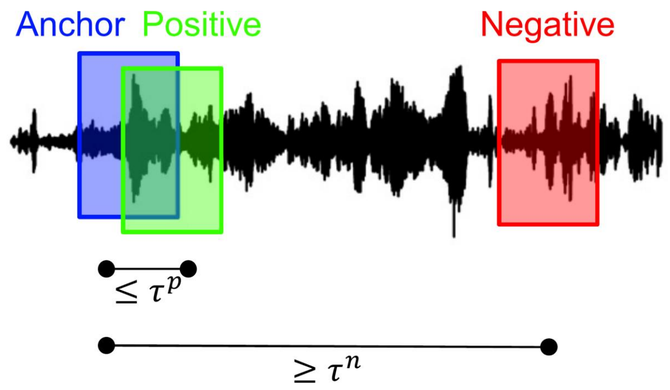
- **Learn from correspondences between different versions of a music piece**
- **Learned features capture instrumentation + outperform a single-version baseline**

1. Representation Learning for Music

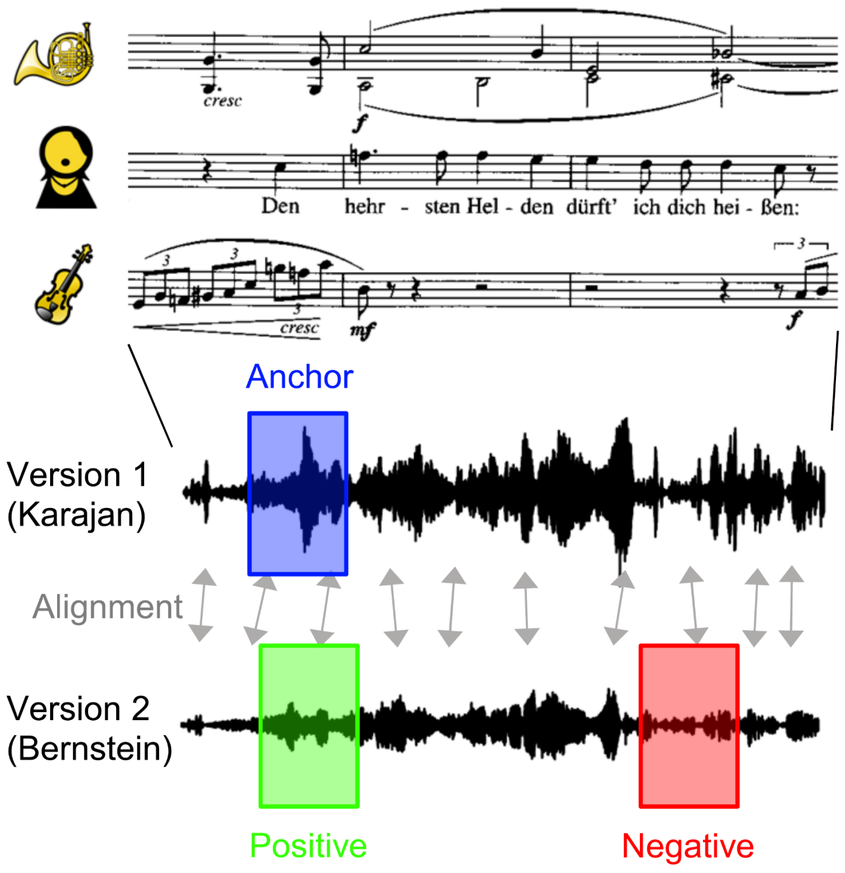
- Learn representations using pretext task
- Apply for downstream tasks



- Pretext task: Learning from temporal proximity



2. Proposed Approach

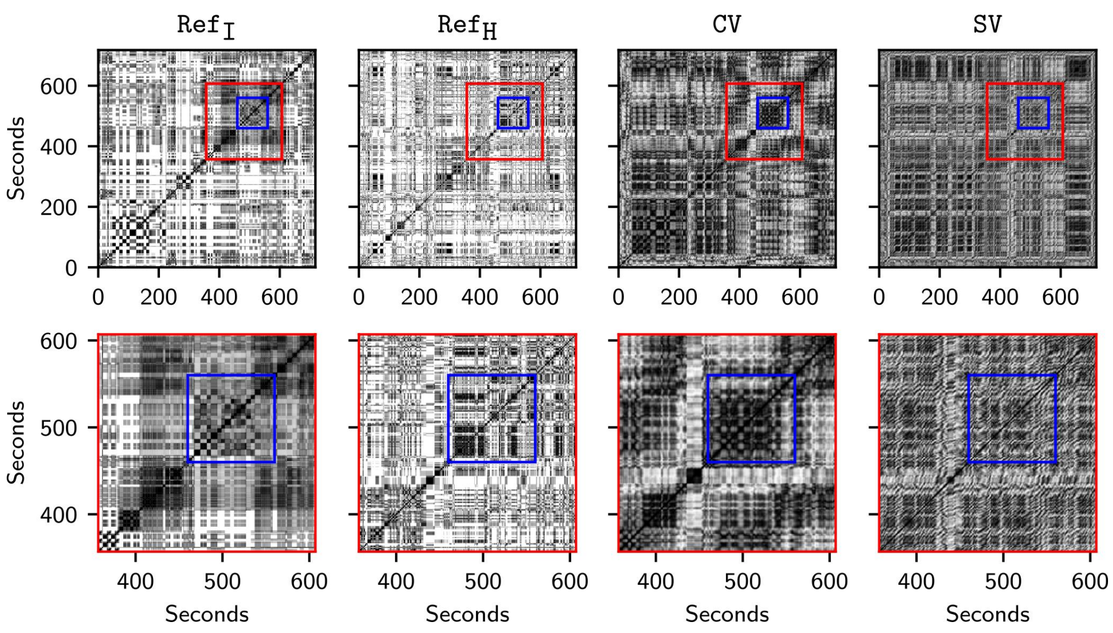


- Key idea: utilize cross-version data
- Score is the same for all versions: same instrumentation / same pitches

3. Evaluation: Setup

- 20h cross-version dataset of orchestral music
- VGG-like CNN architecture
- `CV`: proposed cross-version approach
- `SV`: traditional single-version method
- $\texttt{Ref}_{\texttt{I}}$: reference annotations of instrument activity
- $\texttt{Ref}_{\texttt{H}}$: reference annotations of pitch classes
- `Sup`: supervised instrument classific. baseline

4. Evaluation: Self-Similarity



5. Evaluation: Probing

- Train and evaluate small downstream network for instrument classification

	AP	AUC	F1
SV	0.708	0.735	0.590
CV	0.753	0.795	0.657
Sup	0.838	0.881	0.772

6. Conclusions

- Cross-version pretext task learns instrumentation despite using no instrument labels
- Future work: explore impact of augmentations

References

[1] Matthew C. McCallum, "Unsupervised learning of deep features for music segmentation," in ICASSP, 2019.

[2] Janne Spijkervet and John Ashley Burgoyne, "Contrastive learning of musical representations," in ISMIR, 2021.