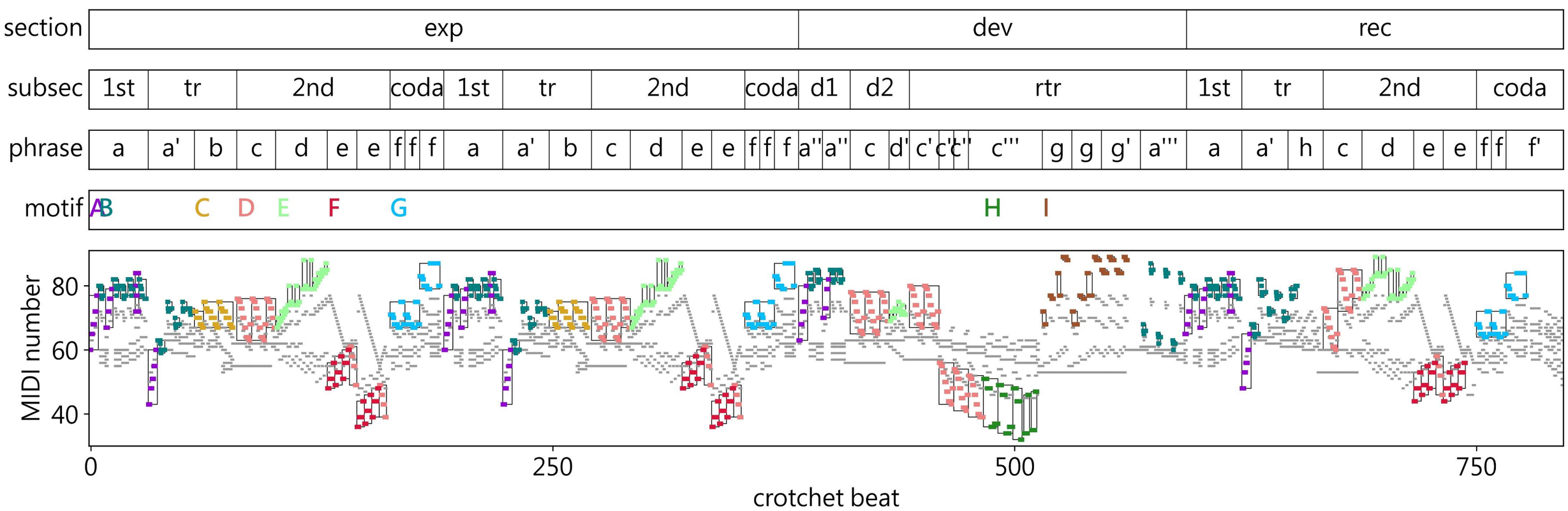
BPS-Motif:

A Dataset for Repeated Pattern Discovery of Polyphonic Symbolic Music

Yo-Wei Hsiao, Tzu-Yun Hung, Tsung-Ping Chen, and Li Su*
Institute of Information Science, Academia Sinica, Taiwan

BPS-motif is a new symbolic music dataset containing the note-level annotation of motives and occurrences in the first movements of Beethoven's Piano Sonatas (BPS).

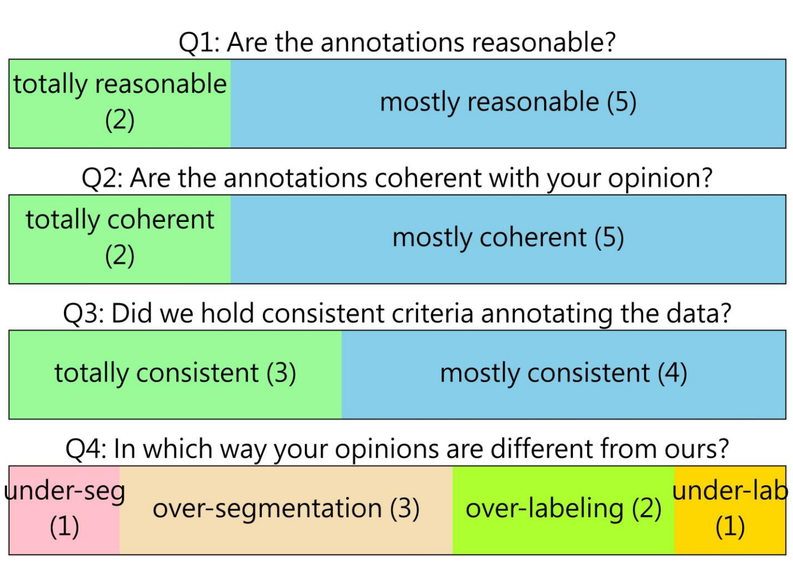


Background

- Repetition: ubiquitous & multi-fold existence in music
- A pattern: a group of notes that
  1. **Is musically important**
  2. **Occurs multiple times in music**.
- Challenges of pattern discovery:
  1. High computational complexity
  2. Tend to find redundant and insignificant patterns
  3. A large pattern can be potentially divided into small ones
  4. Subjectivity regarding repetition, similarity, and musical importance
- Our contribution: **new dataset** and **new motif discovery algorithm**
  1. **Thematic pattern annotation** (i.e., themes, phrases, sections): provided in the BPS-FH dataset (Chen et al., 2018)
  2. BPS-motif: the largest dataset with **motivic pattern annotation** for intra-opus motif discovery

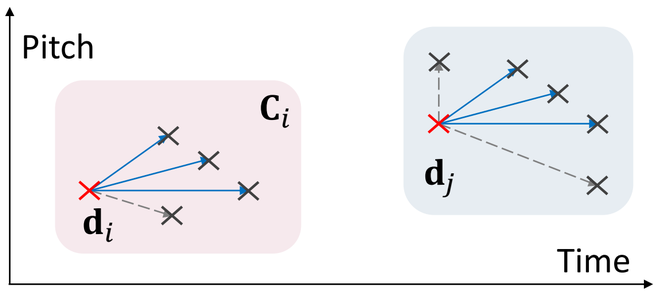
Proposed dataset

- **32 first movements**
- **263 distinct motives** with **4,944 occurrences** in total
- **36,652 motif notes** (28.87% of the total number of 126,943 notes)
- Average 8.22 motives and 154.5 occurrences per music piece
- Average 7.41 notes per motif
- Annotation process: **2 annotators** and **7 external reviewers**, with a **review** and **discussion phase**
- Responses to review questions:



Motif discovery algorithm

- A music piece $\mathbf{D} \coloneqq \{\mathbf{d}_i\}_{i=1}^{N}$, $\mathbf{d}_i \coloneqq (o_i, p_i)$ denotes the ith note, and o_i, p_i denote its onset and pitch value
- **SIATEC algorithm** (Meredith et al., 2002): finds all the translational equivalence class (TEC) of the available maximal translatable patterns (MTPs) in $\mathbf{D}$: $\mathrm{MTP}(\mathbf{v}, \mathbf{D}) \coloneqq \max_{|\mathbf{d}|}\{\mathbf{d}: \mathbf{d} \in \mathbf{D} \text{ and } \mathbf{d} + \mathbf{v} \in \mathbf{D}\}$
- **Common structure** (proposed):
  1. A segment as a set of vectors
  2. The common structure of two segments: the vectors which exist in both segments (blue vectors)
  3. The most common structure for some $\mathbf{d}_j$ w.r.t $\mathbf{d}_i$: two occurrences



Data	Algorithm	P(est)	R(est)	F(est)	P(occ)	R(occ)	F(occ)	P(thr)	R(thr)	F(thr)	Runtime (s)
BPS-motif	SIATEC	0.1804	0.6444	0.2803	**0.2102**	0.2771	**0.2235**	0.0408	**0.2994**	0.0713	**28.5082**
	Proposed	**0.5709**	**0.8339**	**0.6733**	0.1491	**0.4174**	0.2002	**0.1222**	0.2644	**0.1646**	119.533
JKU-PDD	SIATEC	0.1238	0.4630	0.1920	**0.5248**	0.3970	**0.4437**	0.0706	**0.4006**	0.1176	**1.5099**
	Proposed	**0.2649**	**0.5002**	**0.3406**	0.4208	**0.5105**	0.3948	**0.1096**	0.3003	**0.1561**	4.0546