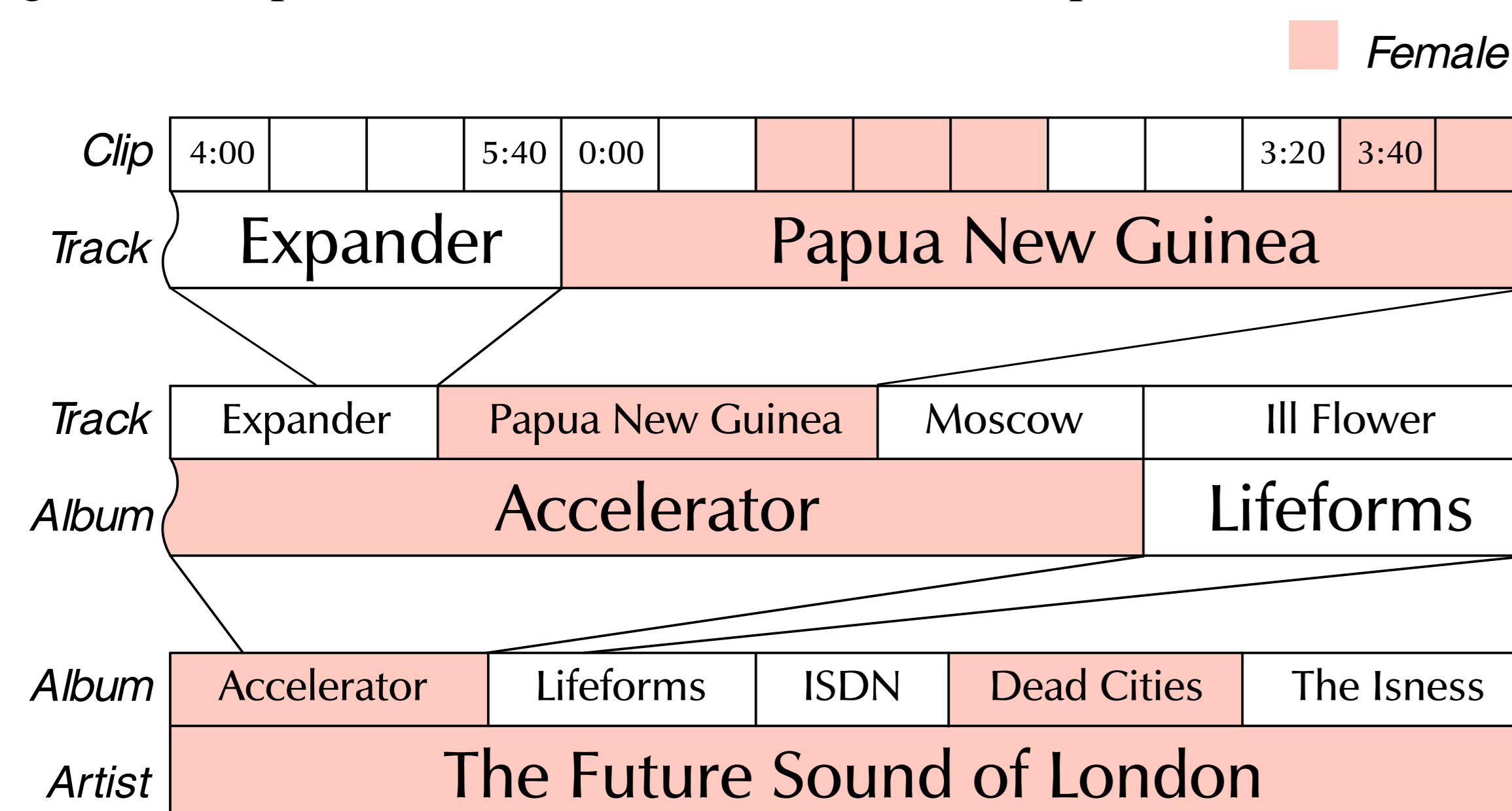


## 1. Summary

- Multiple-instance learning (MIL) algorithms train classifiers from lightly supervised data
  - collections of instances, called bags, are labeled rather than the instances themselves
  - algorithms can classify bags or instances, we focus on instances
- Motivation for applying MIL to MIR:
  - propagate metadata between granularities: artist, album, track, 10-second clip
  - e.g. train clip classifiers using metadata from Last.fm, Pandora, the All Music Guide, etc.
- This work compares 2 MIL algorithms, mi-SVM and MILES, on tag classification
  - Data: tags at track, album, and artist granularities, derived from MajorMiner clip tags
  - Two tasks: recover tags for training clips, predict tags for held-out test clips.
  - Results: mi-SVM better than control at recovery, comparable at prediction
  - MILES performs adequately on recovery task, but poorly on prediction task

## 2. Metadata granularity

- Instances: the atomic elements being classified, in this case 10-second clips of songs
- Bags: labeled collections of instances, in this case tracks, albums, and artists
  - A bag is labeled negative if *no* instance in it is positive
  - A bag is labeled positive if *at least one* instance in it is positive



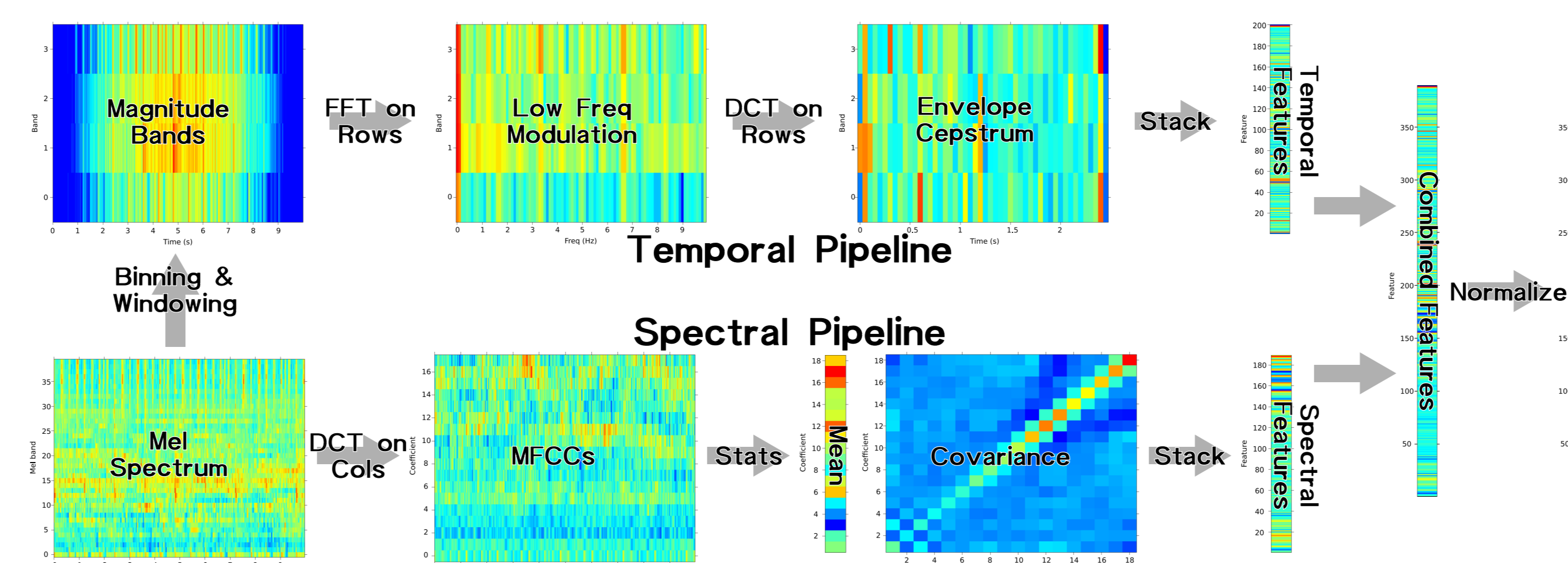
## 3. MajorMiner data for multiple-instance learning

- Data collected by the MajorMiner game (Mandel and Ellis, 2007)
  - players label 10-second clips with arbitrary textual descriptions called *tags*
  - score points when others describe the same clips with the same tag
  - these experiments include tags verified by  $\geq 2$  players on  $\geq 35$  clips
  - 43 tags qualify, total of 9000 verifications on 2200 clips
- These data are well suited to multiple-instance learning experiments
  - most MIL datasets labeled at bag level, difficult to evaluate at instance level
  - MajorMiner tags are applied directly to clips and derived for bags
  - can test instance-level classification in both the training set and a separate test set
- One drawback is that these data do not include negative labels

## 4. Multiple-instance learning algorithms

- mi-SVM (Andrews et al., 2003)
  - Assume bag labels apply to all instances
  - Iterate until labels no longer change:
    - train SVM on current labels
    - use SVM to impute labels for examples in positive bags
- MILES (Chen et al., 2006)
  - Create “feature” matrix for bags, distance between bag and every instance
  - Use 1-norm SVM to simultaneously select features and learn separating hyperplane
  - Infer instance classifications from bag classifications, positive, negative, or indifferent
- Naïve SVM: assume bag labels apply to all instances, train SVM on those labels
- Cheating SVM: train SVM on true instance labels, upper bound on recovery accuracy

## 5. Spectral and temporal feature extraction



## 7. Conclusions

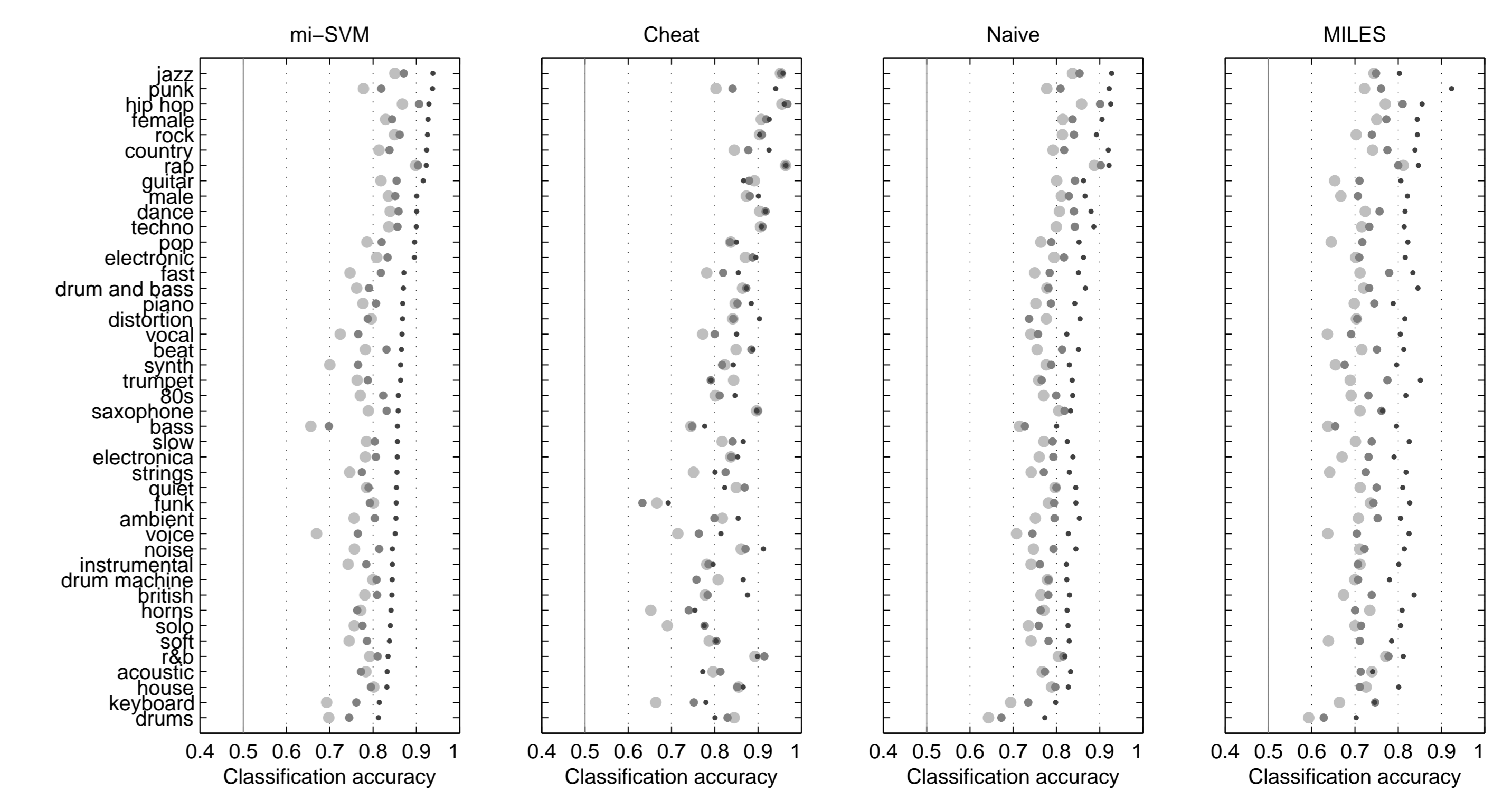
- The granularity of music metadata makes multiple-instance learning necessary for MIR
- Classifiers are more accurate at recovery than prediction because of partial supervision
- Classifiers are more accurate using finer bag granularities than coarser
- mi-SVM is more accurate using track bags than artist bags for most “small-scale” tags  
saxophone, synth, piano, soft, vocal, not trumpet
- MILES’ failure at prediction possibly caused by differences in bag size, incorrect bias
- Cheating control is less accurate than naïve control on prediction task, surprisingly
- Many other potential applications of MIL in MIR: polyphonic transcription, singing voice detection, structure finding, and instrument identification

## References

- S. Andrews, I. Tschantzaris, and T. Hofmann. Support vector machines for multiple-instance learning. In S. Thrun and K. Obermayer, editors, *Advances in Neural Information Processing Systems 15*, pages 561–568. MIT Press, Cambridge, MA, 2003.
- Y. Chen, J. Bi, and J. Z. Wang. MILES: Multiple-instance learning via embedded instance selection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(12):1931–1947, 2006.
- M. I. Mandel and D. P. W. Ellis. Song-level features and support vector machines for music classification. In J. D. Reiss and G. A. Wiggins, editors, *Proc. Intl. Symp. Music Information Retrieval*, pages 594–599, September 2005.
- M. I. Mandel and D. P. W. Ellis. A web-based game for collecting music metadata. In S. Dixon, D. Bainbridge, and R. Typke, editors, *Proc. Intl. Symp. Music Information Retrieval*, pages 365–366, 2007.

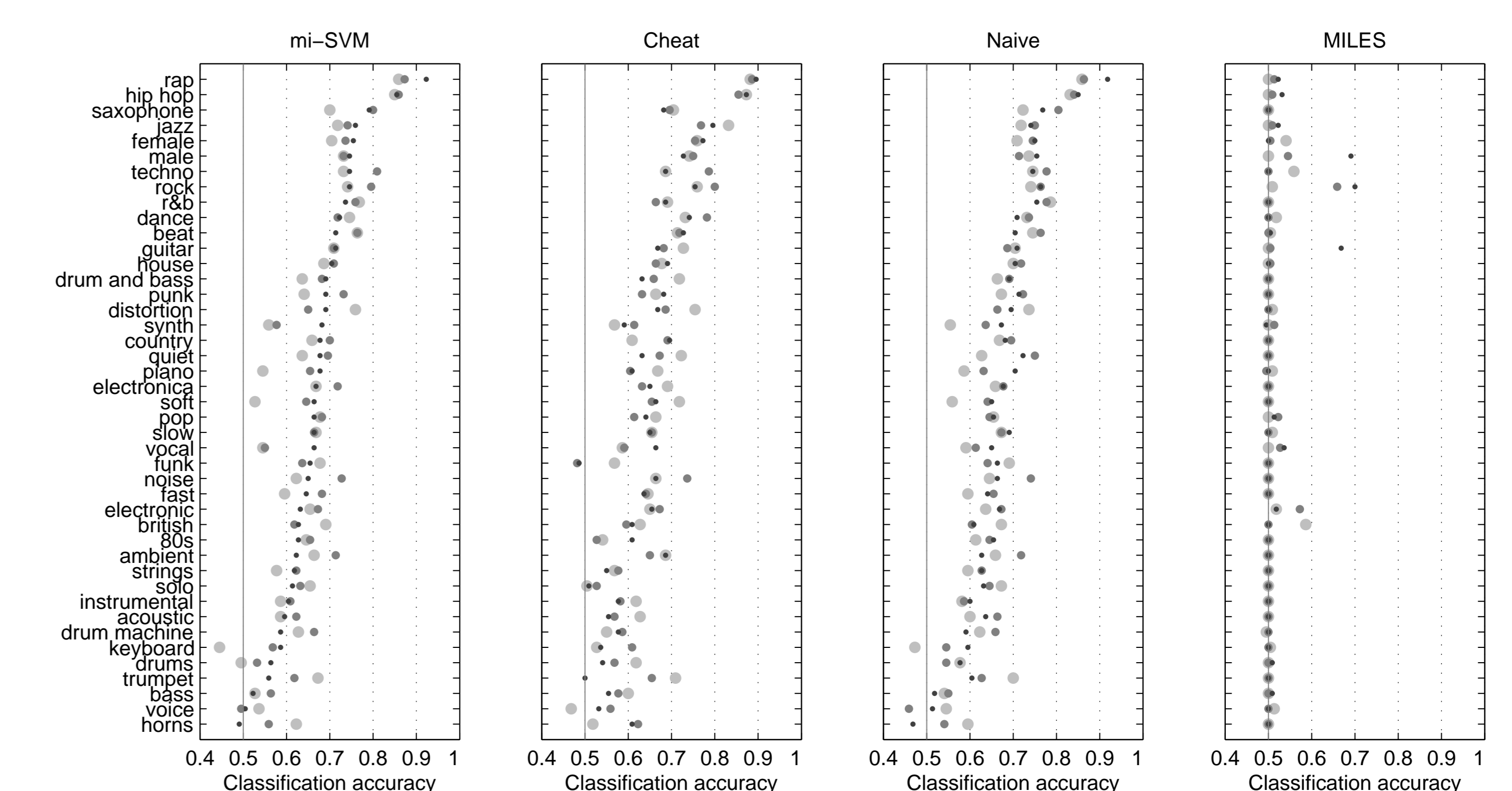
## 6. Multiple-instance autotagging experiments

- 2-fold cross validation, artist filtering, 5 different splits
- Each tag treated as a separate binary classification task
- Accuracy measured on balanced positive and negative instances, constant 50% baseline
  - recovery evaluated on as many clips as possible, varied by tag
  - prediction evaluated on 11 positive and 11 negative instances per fold  $\Rightarrow N = 220$
- Bags at the track, album, artist granularities contained on average 2.47, 4.44, and 8.17 clips



Recover tags for training clips, significant difference varies.

Each pane is an algorithm, and each style of dot is a different bag granularity. Dots get larger and lighter for coarser granularities: track, album, artist.



Predict tags for test clips, significant difference around 0.06

|        | Overall      |      |      |          |      |      |
|--------|--------------|------|------|----------|------|------|
|        | Training set |      |      | Test set |      |      |
|        | Trk          | Alb  | Art  | Trk      | Alb  | Art  |
| mi-SVM | 87.0         | 80.9 | 78.0 | 66.8     | 67.8 | 65.4 |
| MILES  | 81.2         | 73.2 | 70.0 | 51.7     | 50.9 | 50.6 |
| Naïve  | 85.0         | 79.5 | 77.3 | 67.4     | 67.7 | 66.0 |
| Cheat  | 85.9         | 83.9 | 82.6 | 64.7     | 65.7 | 66.2 |