Artist Detection in Music with Minnowmatch

Brian Whitman
Gary Flake
Steve Lawrence

Computer Science
NEC Research Institute, Princeton NJ

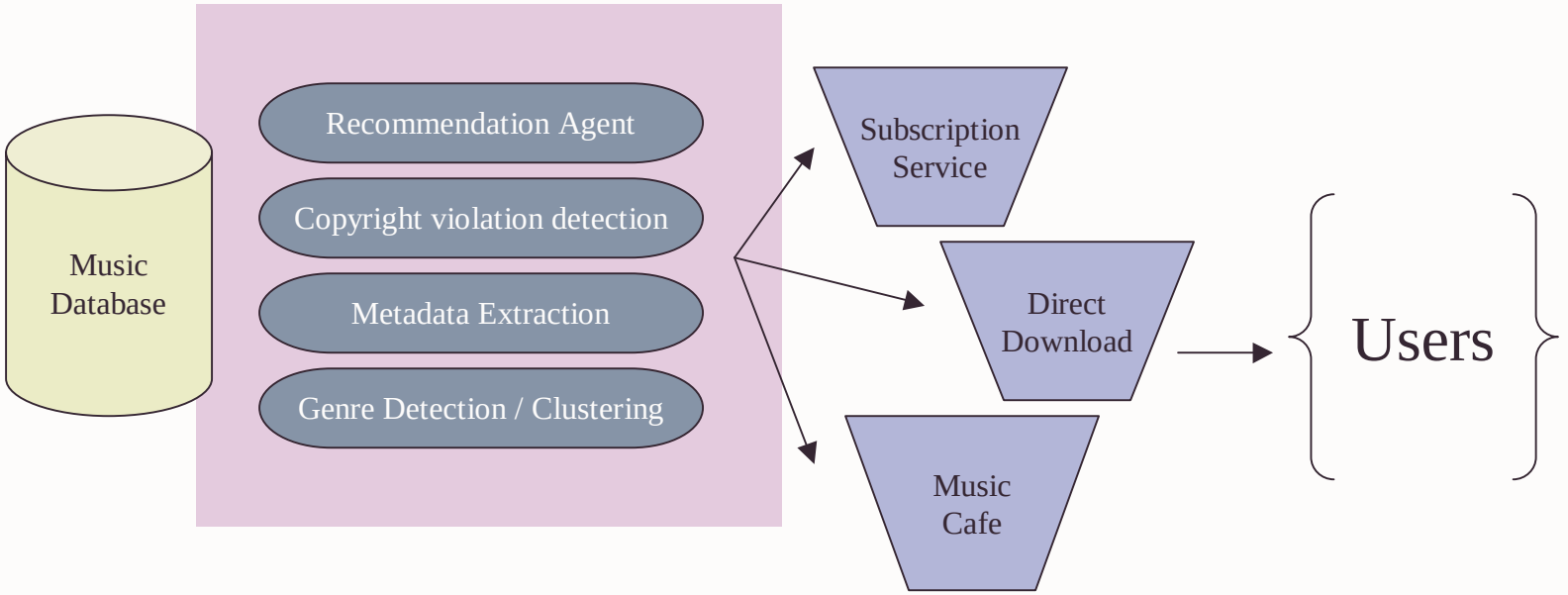
NNSP - 9/2001

Machine Listening / Music IR

- Understanding sound just as a human would
- Combination of machine learning and digital signal processing
- But not just audio—
  - Score level
  - Web crawlers
  - P2P robots
  - User modeling

What Will Happen

- We'll create intelligence between the data and the end-user / consumer:



Artist Recognition

- Formally: creates a representation of source X for future offline classifications
- Like web topic detection, but for music
- Subjective? Human trait?
- Assumptions:
  - No 2 equivalent artists
  - Listeners familiar with similar sounding music can still identify artist

2 Problems

- Feature representation of digital audio for machine learning
- Multi-class learning with support vector machines

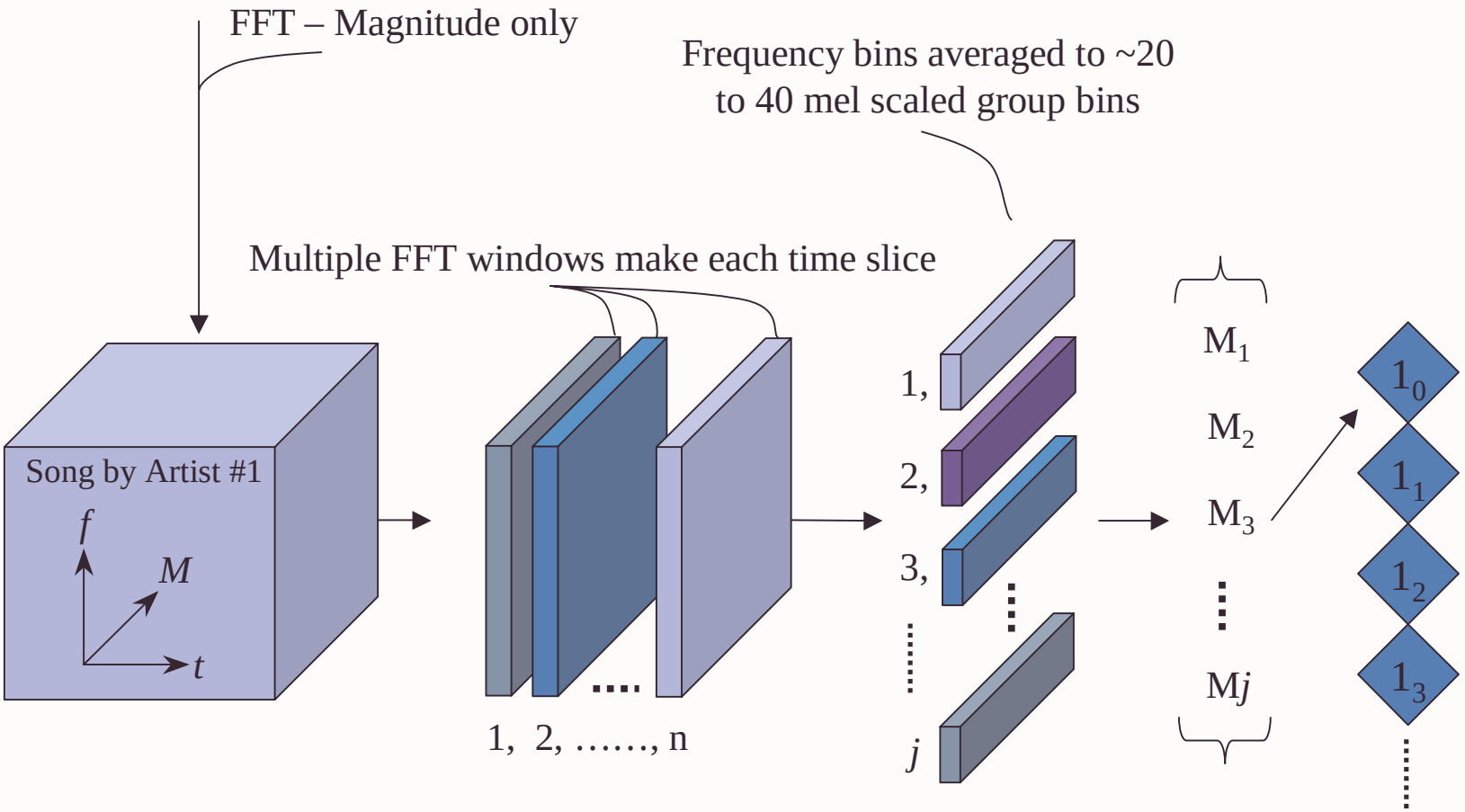
Feature Space

- ‘Perceptual Representation’ (PR) →
  - FFT, MFCC, Wavelets
- Merge adjacent windows into 2-beat long ‘time slices’
- Merge adjacent frequency bins (FFT) into a mel-scaled ‘group bin’
- Each timeslice vector is fed to a learner with time-delay embedding

Short Time Slices

- “Car Seek”
- Enough time to get the “who”
- But not the “what”
  - (Song rec– very different)
- Slices about 2 beats long work best
- Allow for slice level learning decisions

Perceptual to Learning Process



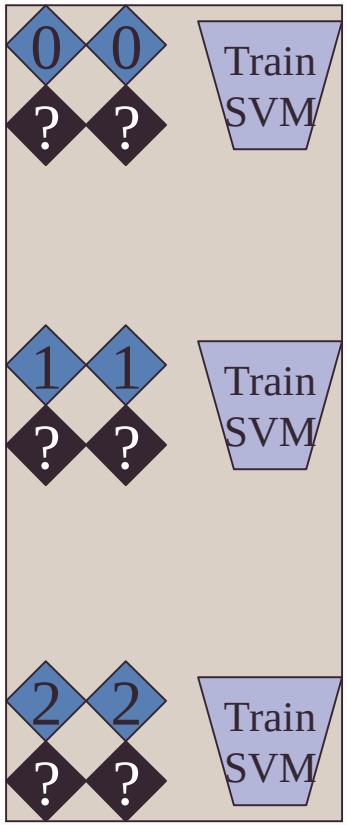
Learning Methods

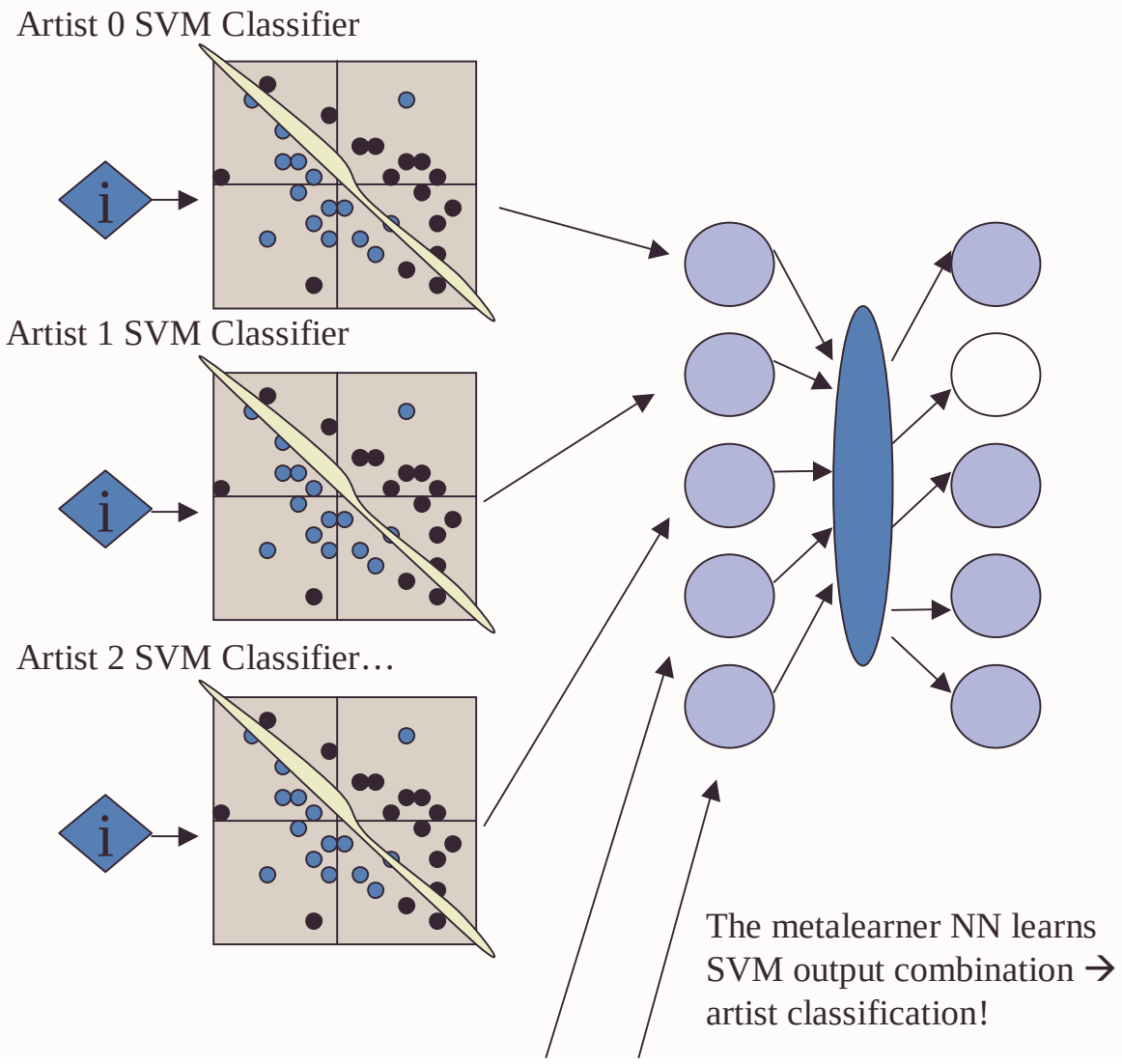
- Multiple methods tried:
  - Simple NN multi-classifier
  - Multiple SVM classifiers with choose-highest
  - SVM / NN metalearner

SVMs – 0/1 classification only

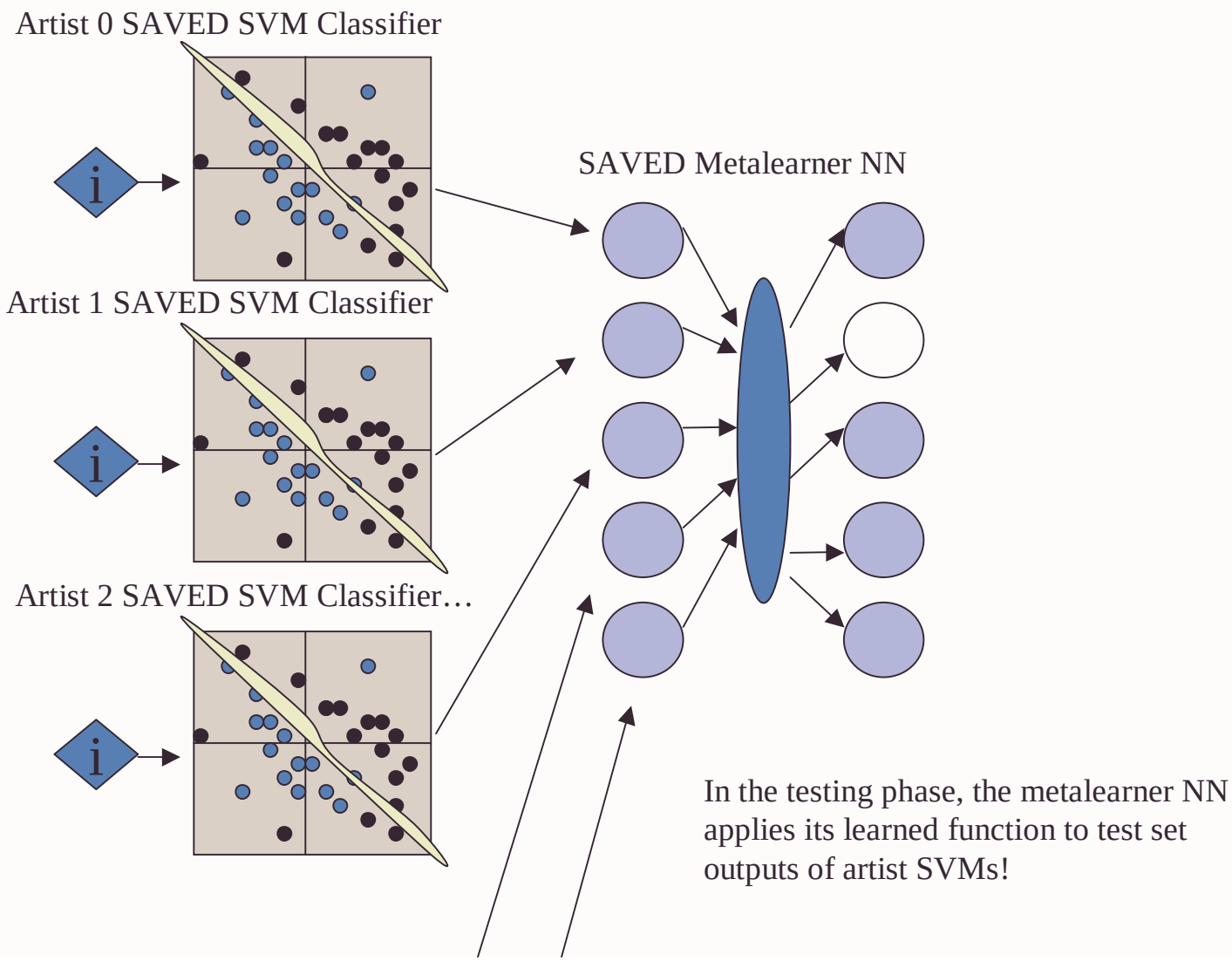
- We analyze the n-dimensional artist space to determine the most likely artist and an associated confidence value
- Previous methods with multiple SVMs:
  - Highest output
  - $(S_{h0} - S_{h1}) * S_{h0} >$ threshold
  - Log scaling
- We can learn this relation instead!

Metalearner Training





Metalearner Testing



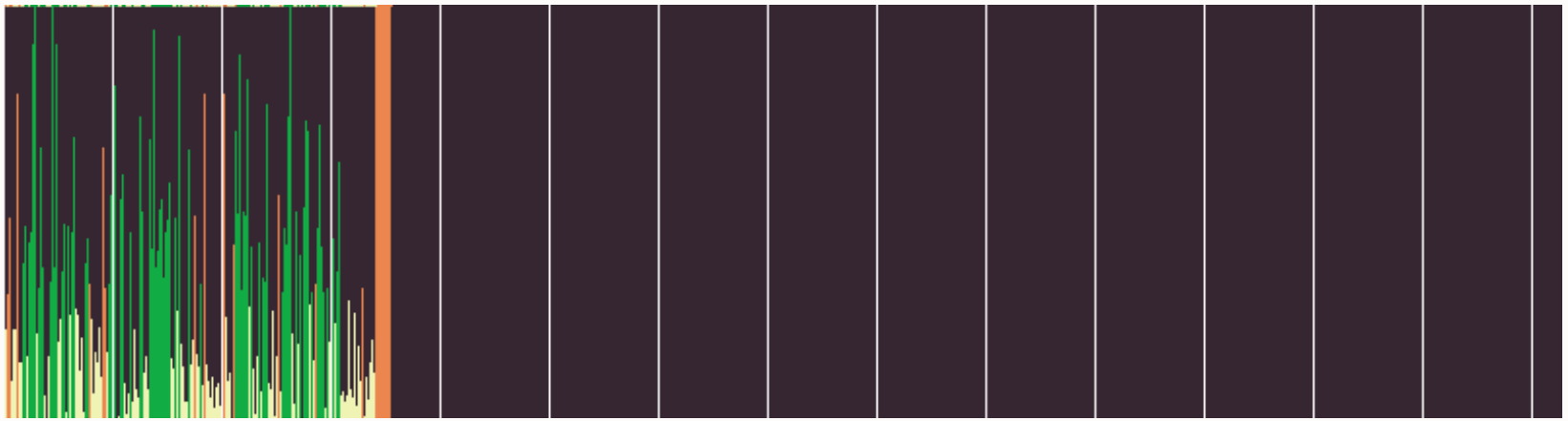
In the testing phase, the metalearner NN applies its learned function to test set outputs of artist SVMs!

Metalearner

- Each time slice is now represented as a set of *artist magnitudes*
- {artist0=.3, artist1=.02, artist2=–23.4}
- Also a great representation for other tasks!
  - Mima's recommender
  - Genre identification

Confidence

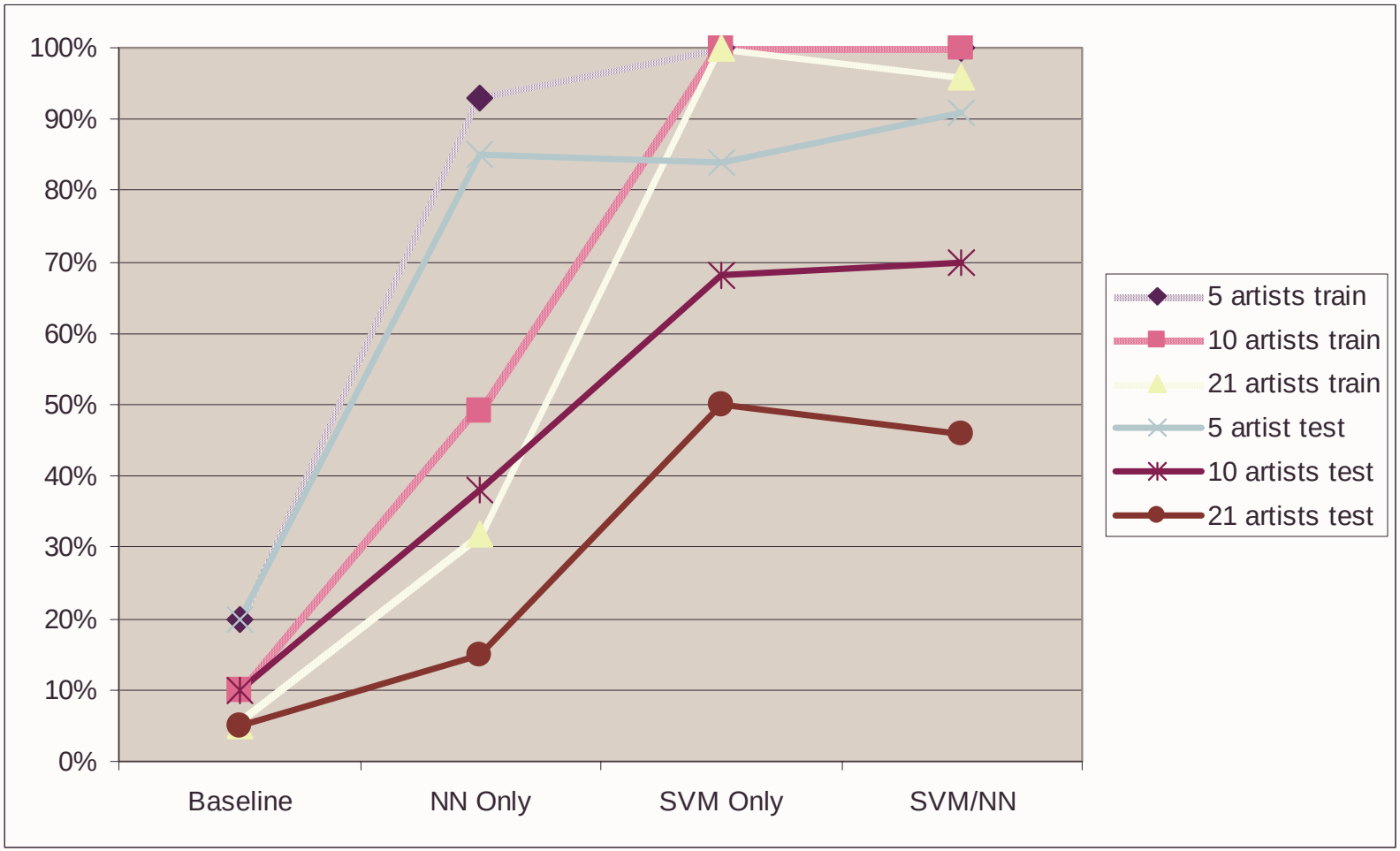
- Allows for slice level decisions
- Can throw out low-energy frames or ‘set-similar’ sound slices



Experiments and Results

- 3 learning methods, different artist set sizes
- “Intentionally Confusing”:
  - Artists from similar genre in each set
  - Song sets have 5-6 songs per “collection” per artist

Results


100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Baseline
NN Only
SVM Only
SVM/NN
5 artists train
10 artists train
21 artists train
5 artist test
10 artists test
21 artists test

Conclusions, Future

- Need better music representations to combat scaling problems
- System could be extended to non-audio topic / aboutness classifications