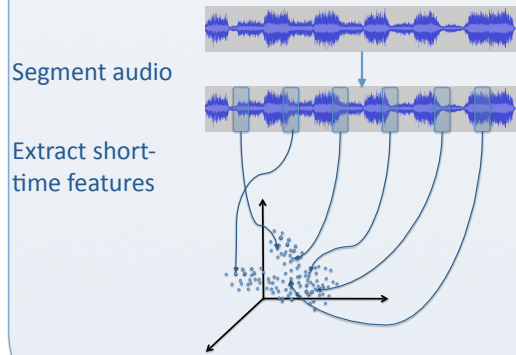# Automatically Annotating Music with Tags

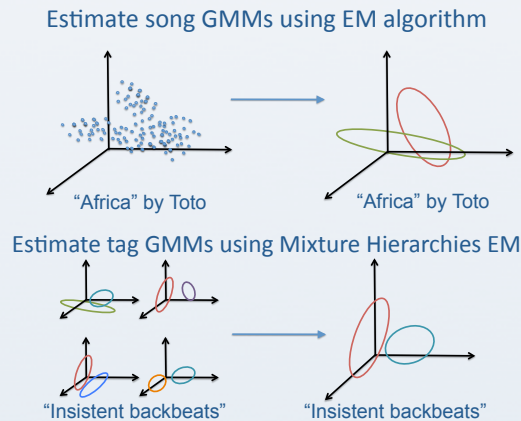Derek Tingle, Youngmoo Kim, Doug Turnbull Swarthmore College, Drexel University

Our goal is to automatically annotate music with text-based tags like "bluegrass roots", "mixed acoustic instrumentation", and "minor key tonality". Once a song has been annotated, we can retrieve it using a semantic music discovery engine (a.k.a. **"Google for Music"**). In this work, we compare two approaches to autotagging: classification using short-time features and classification using song-level features.
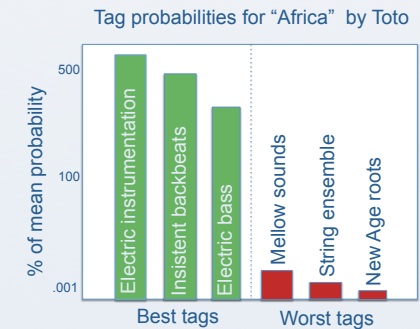
Short-Time Features

Segment audio

Extract short-time features

Model Songs & Tags

Estimate song GMMs using EM algorithm

"Africa" by Toto → "Africa" by Toto

Estimate tag GMMs using Mixture Hierarchies EM

"Insistent backbeats" → "Insistent backbeats"

Tag New Songs

Evaluate the song likelihood for each tag

Tag probabilities for "Africa" by Toto

% of mean probability

Best tags — Electric instrumentation, Insistent backbeats, Electric bass

Worst tags — Mellow sounds, String ensemble, New Age roots

Song-Level Features

Upload audio

the echonest

Download song-level features → Key, Mode, Time Signature, Beats Per Minute, Etc.

Model Tags

Learn a Support Vector Machine (SVM) or Boosted Decision Stump (BDS) to classify tags

Song Loudness (y-axis)

Beats per Minute (x-axis)
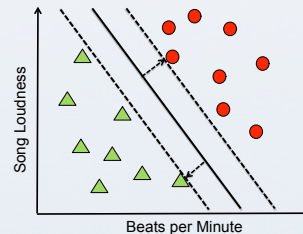
Results

Classifier	AUC
Baseline	0.5007
Short-time features & GMM	0.8407
Song-level feature & BDS	0.7613
Song-level features & SVM w/ linear kernel	0.6131
Song-level features & SVM w/ RBF kernel	0.6862

We find that short-time features perform better than song-level features on most, but not all, of the tags. Future work will attempt to incorporate both short-time and song-level features into one classifier to utilize the benefits of each feature set.