***Outline***

**Transcription of Percussion Instruments in Polyphonic Audio**

Beach Clark, Daniel Ethridge, and Laney Light

**Intro**

* Problem statement: transcribe specific percussive sounds in a polyphonic texture
  + Kick drum, snare, hi hat
* Motivation
* Background: literature

**Data**

* MDB drums
  + Dataset containing 23 tracks and 7994 onsets
  + Annotated with instruments at specific times
  + Using the combined tracks, containing drums and other instruments
* Additional data as needed

**Methods**

* Onset detection
  + *Madmom*: Python audio signal processing library
    - Recurrent neural network (RNN) onset processor
    - Returns 1 vector per file: for each window in a user defined number of milliseconds, 1 if an onset occurred and 0 if not (using window same size as block size (approx. 11 ms) for baseline
* Feature identification
  + 1st 5 MFCCs, RMS, spectral flux, spectral crest, zero crossing, spectral centroid
  + Normalize data
* Feature selection
  + For mid-term: start with all features we have (possibly narrow down once we have more features calculated, using backward selection)
  + End result: matrix with dimensions (n features) x (m windows), includes all 23 files
* Classification
  + Baseline system: Support Vector Machine (SVM)
    - Describe SVM, pros and cons
    - Implemented using scikit-learn
      * Tried different kernels (linear, polynomial, Gaussian, sigmoid)
      * Multi-class categorization (6 percussion sounds)
        + May need to reduce to 4 because 2 are very rare, usually predicts it’s not there
      * SVM can handle all possible permutations
        + Multiple instruments can be at the same time (or close to the same time, timing could be slightly off but in the same x ms window)
        + Downside – hard to understand what exactly the SVM is doing
    - Separate into training and testing datasets (75% training, 25% testing)
* Results
  + Preliminary data analysis
    - N instruments, N occurring at approx. same time
  + Compare to ground truth from MDB drums dataset
    - Best fitting kernel - precision, recall, F1, confusion matrix
    - Compare different kernels – just look at F statistic
  + Visuals of results
    - Heat map, tables
* Next steps
  + Data set is small
    - use N-fold cross validation since we only have 23 files
    - supplement with additional data
  + Remove 2 instruments that don’t occur often
  + Not working well for kick drum
    - Low pass filter
  + Determine which features have biggest impact, possibly narrow down with backward selection
    - Scatter plots
    - logistic regression – different model for each instrument, see which features are important for each instrument
  + Different classification algorithms
    - KNN
* Future work
  + Improve efficiency, look at run times