# Beatbox Classification Using ACE

Elliot Sinyor, Cory McKay, Rebecca Fiebrink, Daniel McEnnis, Ichiro Fujinaga
Music Technology Area, Faculty of Music, McGill University, and CIRMMT, Montreal, Canada

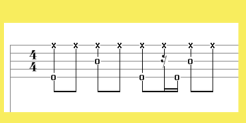
## Project Description

The goal of this project is to identify a reliable way to classify vocal percussion sounds.

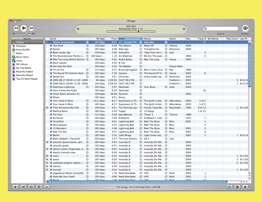
Three expert beatboxers and two non-beatboxers were recorded resulting in a collection of 1242 samples.

ACE (Automatic Classification Engine) as well as a Genetic Algorithm-based feature-selection system were used.

Potential Applications:



Automatic transcription



Query by beatboxing

Control of sound generation

## Background

Beatboxing refers to the mimicking of drum sounds with one's mouth.
The practice emerged in the early 1980's as part of hip-hop culture.

Rahzel

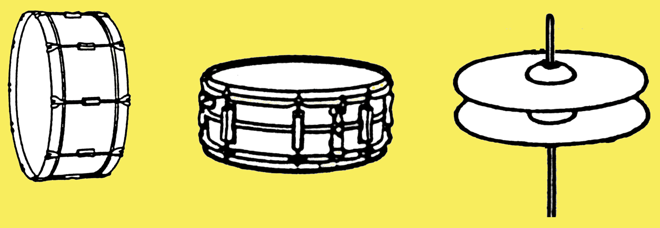
Proficient beatboxers, like Rahzel pictured above, can vocalize bass-lines, melodies and drum sounds simultaneously.

This project focuses solely on drum sounds, specifically those based on unvoiced plosives and fricatives such as /p/, /t/, /k/, and /s/.

## Recent Efforts

Kapur, Benning, Tzanetakis (2004)

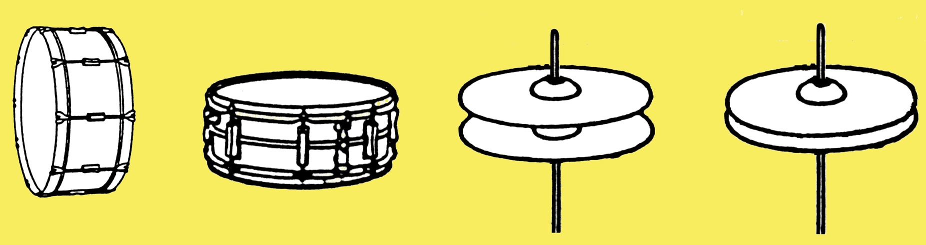
Among a number of related experiments, the authors attempted to classify:



75 samples, 3 classes, 1 feature (zero-crossing rate) → 97.3% accuracy

Hazan (2005)

For the purpose of automatic transcription, the author attempted to classify:



242 samples, 4 classes, 28 features → 86% accuracy

## Our Approach

The following setup was used to record our 5 subjects:

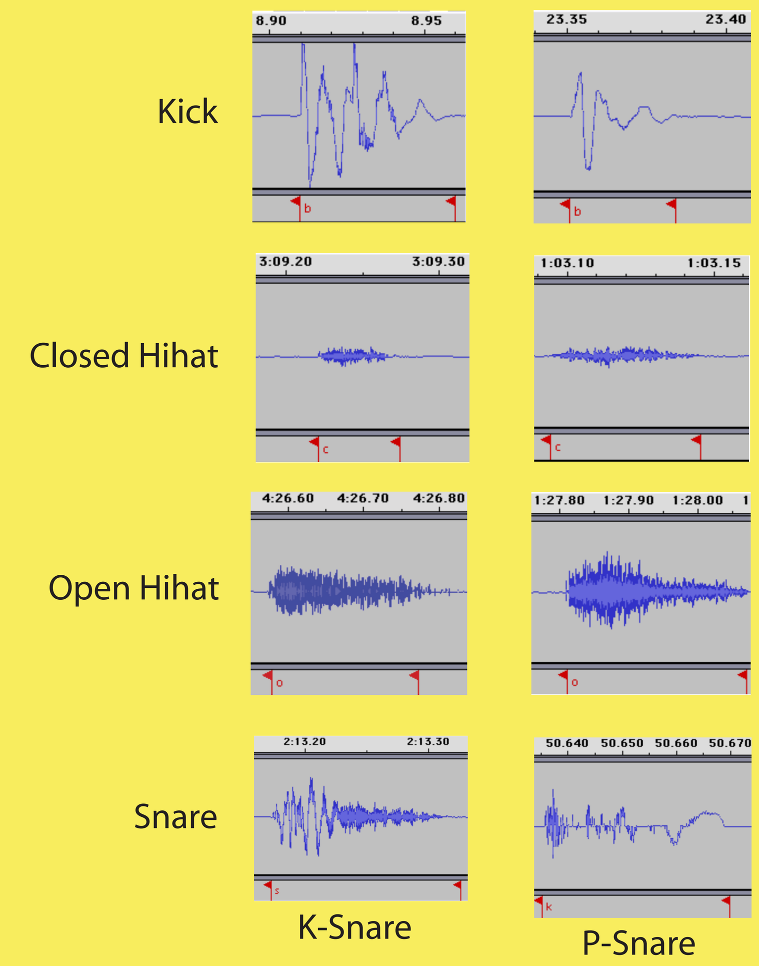

Subjects were told to make kick drum (aka bass drum), snare, open hihat, and closed hihat sounds.

The subjects were found to make two different snare sounds, one resembling a "k" sound and the other resembling a "psss" sound, so we subdivided the snare class into k-snare and p-snare.

## Segmentation

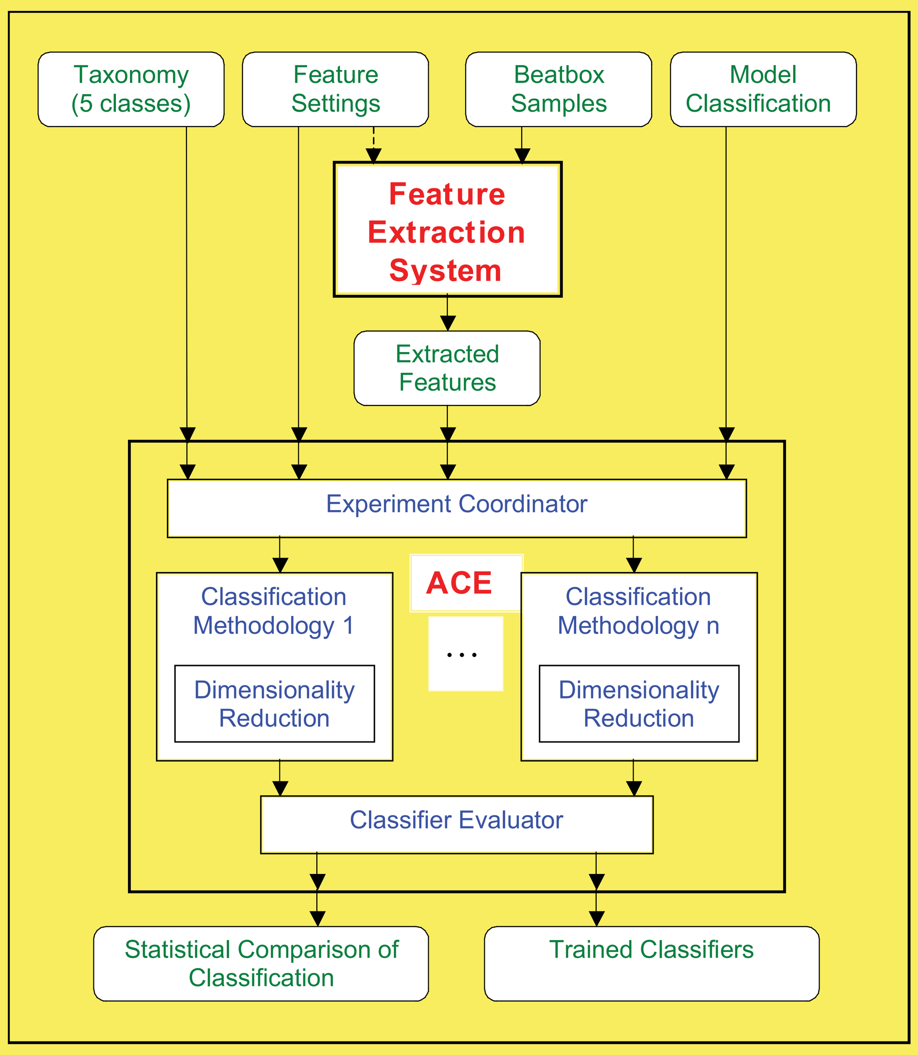
Segmentation was performed manually using Audacity, an open-source, multi-platform audio editor.

The data was then divided into 5 classes:

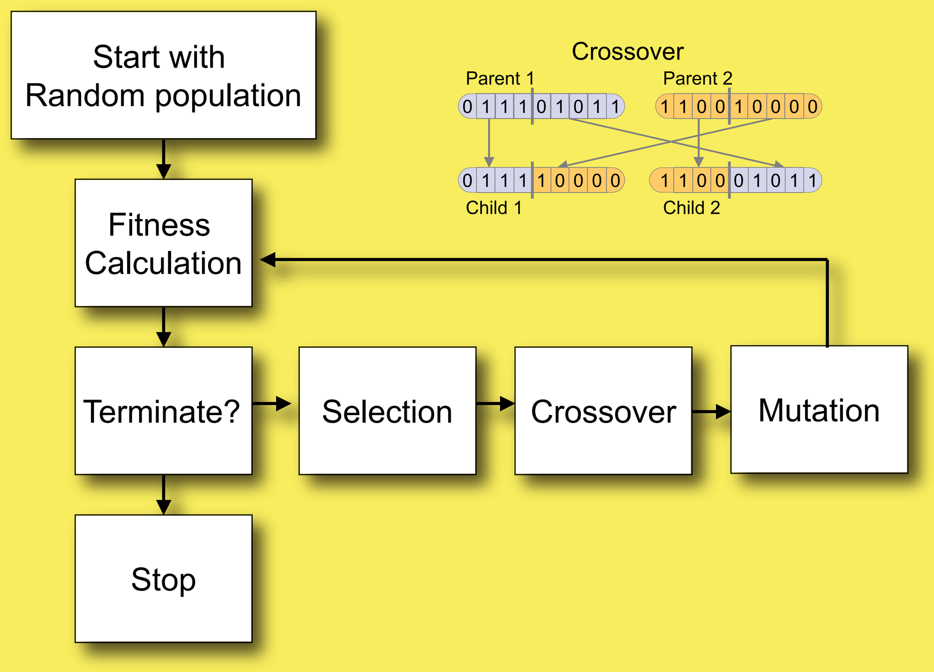



## ACE Classification

The sample set and taxonomy were entered into ACE. The features used are listed in the next section.




## GA-based Feature Selection




In our case, fitness was measured by the classifier accuracy rate for a given set of features.

The following features were used by ACE. Features marked "true" were selected by the GA system.

| | |
|---|---|
| Sp_Centroid_Overall_Avg | false |
| Sp_Centroid_Overall_Std_Dev | false |
| **Sp_Rolloff_Point_Overall_Avg** | **true** |
| **Spectral_Rolloff_Point_Ovl_Std_Dev** | **true** |
| Sp_Flux_Overall_Avg | false |
| Spectral_Flux_Overall_Std_Dev | false |
| **Compactness_Overall_Avg** | **true** |
| **Compactness_Overall_Std_Dev** | **true** |
| **Spectral_Variability_Overall_Avg** | **true** |
| **Spectral_Variability_Overall_Std_Dev** | **true** |
| RMS_Overall_Avg | false |
| **RMS_Overall_Std_Dev** | **true** |
| **RMS_Derivative_Overall_Avg** | **true** |
| **RMS_Derivative_Overall_Std_Dev** | **true** |
| **ZC_Overall_Avg** | **true** |
| **ZC_Overall_Std_Dev** | **true** |
| **ZC_Derivative_Overall_Avg** | **true** |
| ZC_Derivative_Overall_Std_Dev | false |
| Strongest_Freq_Via_ZC_Overall_Avg | false |
| Strongest_Freq_Via_ZC_Overall_Std_Dev | false |
| **Strongest_Freq_Via_SC_Overall_Avg** | **true** |
| Strongest_Freq_Via_SC_Overall_Std_Dev | false |
| Strongest_Freq_Via_FFT_Max_Overall_Avg | false |
| **Strongest_Freq_Via_FFT_Max_Ov_Std_Dev** | **true** |

## Results

ACE achieved an accuracy rate of 95.55% using AdaBoost with C4.5 decision trees as base learners.

The GA feature-selection system achieved an accuracy rate of 94.55% using a 1-NN classifier and the 14 features listed above. This can be compared to a rate of 89.36% when ACE used 1-NN for all 24 features.

Classified as:

| a | b | c | d | e | Actual: |
|---|---|---|---|---|---|
| 309 | 0 | 1 | 0 | 1 | a = kick |
| 0 | 278 | 12 | 0 | 0 | b = open |
| 0 | 15 | 273 | 6 | 4 | c = closed |
| 0 | 1 | 5 | 149 | 1 | d = k-snare |
| 2 | 1 | 1 | 3 | 130 | e = p-snare |

Confusion Matrix

| Author | # of classes | # of samples | # of feat. | Classifier | Acc. |
|---|---|---|---|---|---|
| Kapur | 3 | 75 | 1 | ANN | 97.3% |
| Hazan | 4 | 242 | 28 | C4.5 w/ boosting | 86% |
| Sinyor | 5 | 1192 | 24 | C4.5 w/ boosting | 95.55% |
| Sinyor | 3 | 1192 | 24 | C4.5 w/ boosting | 98.15% |

Comparison with other recent attempts

## Acknowledgements

The authors would like to thank beatboxers Benjamin Hammond, Jason Levine and Kweku Sam Kwofie as well as non-beatboxers Ansel Brandt and Joseph Malloch for their time and beats.

McGill
CIRMMT Centre for Interdisciplinary Research in Music Media and Technology