[[1]](#footnote-1)

Automatic Recognition of Beatboxing Sounds using a Hidden Markov Model

Ankoor Apte and Sidney Cozier

*Abstract*— Beatboxing or vocal percussion is a form of music that, despite its prominence, has not been studied extensively in Music Information Retrieval (MIR) research. This paper proposes a solution to the problem of automatic transcription (a fundamental MIR task) of beatboxing. T

# Introduction

## A. Motivation

Beatboxing can be defined as the art of vocal percussion, or mimicking real drum machines with the human mouth and voice. Beatboxing performances can be done solo or with a group of singers, like in a cappella music. Due to their dependence on physical features (how the human mouth is shaped) and subjective interpretation (how a drum sound is translated to a vocal sound by each beatboxer), beatboxing sounds tend to vary significantly across individual performers. Despite this, most beatboxing sounds across performers can be loosely recognized as standard drum sounds (e.g. kicks, snares, hi-hats, among others). This is especially true for individuals that are familiar with these standard sounds.

The research presented in this paper could lead to tools for beatboxers to record and transcribe other beatboxers’ performances.

## B. Problem Statement

Given an audio recording of a solo beatboxing performance, design a system that recognizes the percussion sounds (using a learned set of 10 percussion sounds) and identifies their onset locations in the recording.

# System Design

## A. Hidden Markov Model

The proposed approach for beatboxing sound recognition is a Hidden Markov Model (HMM). The overall system is shown below.

The HMM design includes observations, states and a model. In this case, the observations are sections of the input audio signal (or query) that capture each individual sound. There are 10 distinct states that correspond to the 10 pre-defined set of percussion sounds. Finally, the model is a multivariate normal probability distribution that can be used to estimate the hidden state.

## B. Percussion Dictionary

The HMM has 10 distinct states that correspond to sounds in the percussion dictionary. The percussion dictionary was defined from a set of typical Western drum sounds, using Standard Beatboxing Notation (SBN).

1. Percussion Dictionary

| Percussion Sound | Standard Beatboxing Notation |
| --- | --- |
| Kick | B |
| Snare 1 | Pf |
| Snare 2 | Pch |
| Snare 3 | K |
| Snare 4 | ^Ksht |
| Snare roll | rrh |
| Closed hi-hat | t |
| Open hi-hat/crash/cymbal | ts |
| Rimshot | k |
| Lip Oscillation | BB |

## C. Obtaining Observations

Given the input audio query, the system obtains an array of observations and calculates feature

## D. Training Model

The template is used to format your paper and style the text.

## E. Estimating States

The template is used to format your paper and style the text.

## F. Key Assumptions

# Data Collection

## A. Recording Procedure

The

## B. Status

# Results

Before you begin to format your paper, first write and save the content as a separate text file. Keep your text and graphic files separate until after the text has been formatted and styled.

## A. Evaluation Metric

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract.

## B. Results

* Use either SI (MKS) or CGS as primary units. (SI units are encouraged.) English units may be used as secondary units (in parentheses). An exception would be the use of English units as identifiers in trade, such as “3.5-inch disk drive”.
* Avoid combining SI and CGS units, such as current in amperes and magnetic field in oersteds. This often leads to confusion because equations do not balance dimensionally. If you must use mixed units, clearly state the units for each quantity that you use in an equation.

# Discussion

After the text edit has been completed, the paper is ready for the template. Duplicate the template file by using the Save As command, and use the naming convention prescribed by your conference for the name of your paper.

## A. Plots (if we need them for further analysis

The template is designed so that author affiliations are not repeated each time for multiple authors of the same affiliation.

## B. Limitations

Headings, or heads, are organizational devices that guide the reader through your paper.

## C. Future work

Figure Labels: Use 8 point Times New Roman for Figure labels.

# Conclusion

A conclusion section is not required. Although a conclusion may review the main points of the paper, do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions.

Appendix

Appendixes should appear before the acknowledgment.

Acknowledgment

The team thanks Professor TJ Tsai for his guidance throughout the semester and during the project. The team also acknowledges the time and effort put in by the six beatboxers from the Claremont Colleges a cappella groups to record the test data.

References

1. G. O. Young, “Synthetic structure of industrial plastics (Book style with paper title and editor),” in *Plastics*, 2nd ed. vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15–64.
2. W.-K. Chen, *Linear Networks and Systems* (Book style)*.* Belmont, CA: Wadsworth, 1993, pp. 123–135.

1. \*Research supported by ABC Foundation.

   F. A. Author is with the National Institute of Standards and Technology, Boulder, CO 80305 USA (corresponding author to provide phone: 303-555-5555; fax: 303-555-5555; e-mail: author@ boulder.nist.gov).

   S. B. Author, Jr., was with Rice University, Houston, TX 77005 USA. He is now with the Department of Physics, Colorado State University, Fort Collins, CO 80523 USA (e-mail: author@lamar. colostate.edu).

   T. C. Author is with the Electrical Engineering Department, University of Colorado, Boulder, CO 80309 USA, on leave from the National Research Institute for Metals, Tsukuba, Japan (e-mail: author@nrim.go.jp). [↑](#footnote-ref-1)