# Polyphonic Music Transcription using Deep Learning

Aniruddha Zalani (11097) Ayush Mittal (11183)

February 28, 2014

#### Abstract

One of the most important applications in modelling temporal dependencies in high dimentional spaces is Transcription of polyphonic music. In this project we plan to analyse the problem using deep learning methods. We just focus on polyphonic tunes of piano.

#### 1 Introduction

Polyphonic music in piano means that there are two or more independent notes playing on the same time, in contrast to the monophonic music where only one node is played at a time. A lot work has been done on monophonic transcription but the problem of Polyphonic transcription is still open. Many naturally occuring phenomeno have complex sequences that are inherently sequential but the value at next time step cannot be determined only by the knowledge of previous time step. Examples of such phenomeno inculde music, speech, human motion. Most of them are spanned over high dimensional spaces. Word notes appear together in correlated patterns so this affects the conditional probablity.

### 2 Related Work

Some intersting work has been done using non negative matrix factorization techniques [1] and [2]. However, much of the recent work for modelling temporal dependecies has been done using deep learning [3] and [4]. Nicholas et al., [3] used RNN-RBM based model that learns recursive and probabilistic rules from polyphonic music scores of varying complexity. They exploited the ability of RBMs to represent a complicated distribution for each time step. Juhan et al., [4] used deep belief networks and improved state of the art results by jointly training classifiers for multiple notes and also used the learned feature representations for note classifiers. The DBNs are applied to audio spectrograms. For post processing they used HMMs

## 3 Our approach and Dataset

On similar lines we plan to use DBNs pretrained with RBMs for learning temporal dependecies and later we will be doing multiple node training using some linear classifier to identify the notes. Pre-processing will be a cruicial part. We can compute spectogram for a fixed time interval and with some part of the interval overlapping with adjacent time interval. Pre-training is an important aspect with the DBNs. If the nodes are being initialized to some random values then during gradient decent or backtracking instead of global minima, a local minima will be achieved and thus hampering our results. We will be using standard **Piano-midi.de** database [5].

# References

- [1] Arnaud , Arshia et al. "Real-Time Detection of Overlapping Sound Events with Non-Negative Matrix Factorization"
- [2] Paris and Judith Non-Negative Matrix Factorization for Polyphonic Music Transcription, IEEE 2003.
- [3] N. Boulanger-Lewandowski, Y. Bengio and P.Vincent, Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," ICML, 2012.
- [4] J. Nam, J. Ngiam and H. Lee, Classication- Based Polyphonic Piano Transcription Approach Using Learned Feature Representations," ISMIR, pp. 175-180, 2011.
- [5] http://www.piano-midi.de/