Computational Music

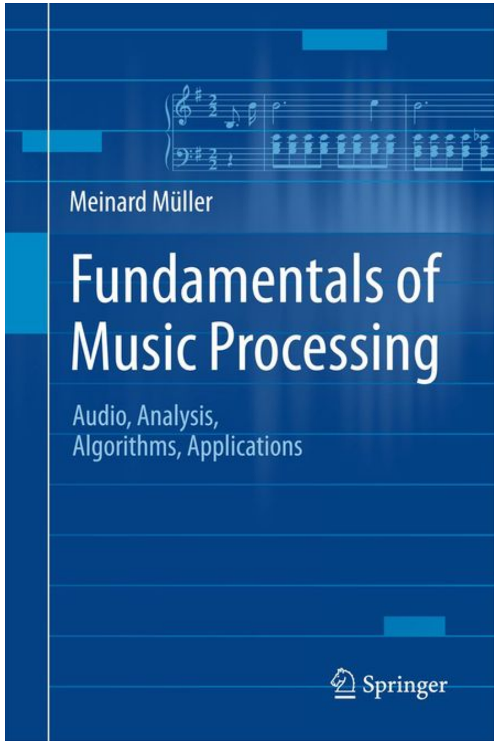
William Steimel

Table of Contents

Musical Data Representations

William Steimel

Source



Chapter 1- Music Representations

Music Representations

- There are three main classes of music representations (Types of Data):
  - Sheet Music
  - Symbolic
  - Audio

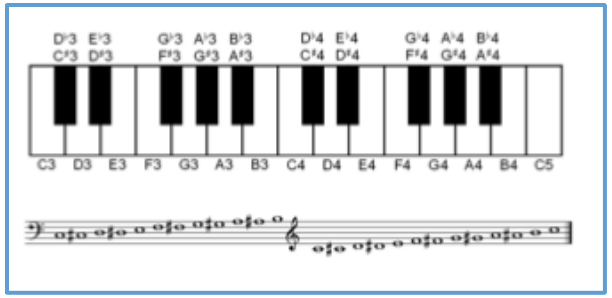
Sheet Music Representations

What is Sheet Music?

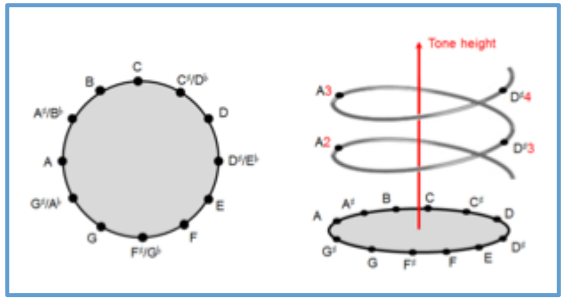
- Visual representations of music given in printed form or digitized images
  - Also known as a musical score
- A formal language for music based on musical symbols and letters
- This book goes very heavily into the basics of reading music which is essential for Music Processing but I will not cover it in this presentation.
  - Summary is contained in next slide

Elements of Musical Sheet Music (Quick Summary)

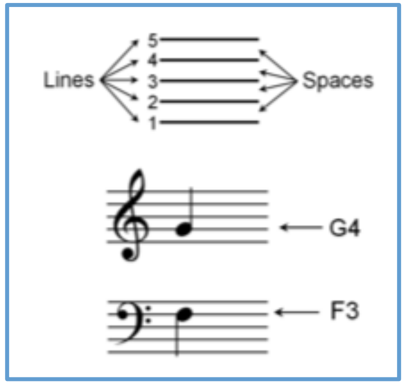
Musical Notes Corresponding to Piano



Chromatic Circle & Shepard's Helix of Pitch

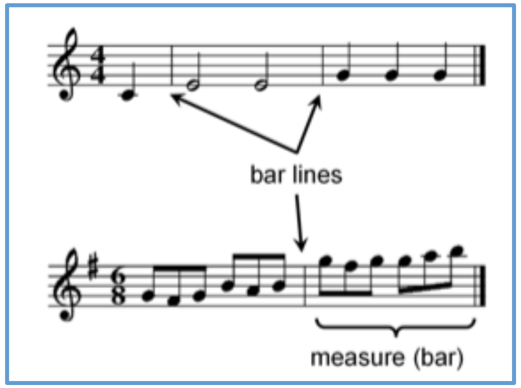


Musical Staff



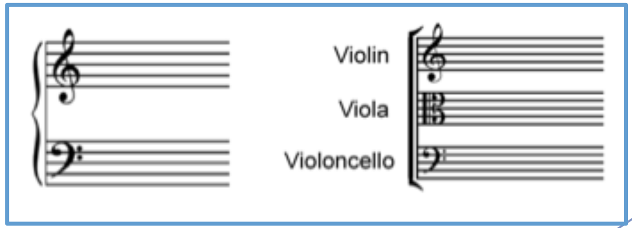
Musical Score with 2 scales

Time Signature & Bar Lines



Note Value & Durations

Piano & Multiple Instruments



Musical Dynamics

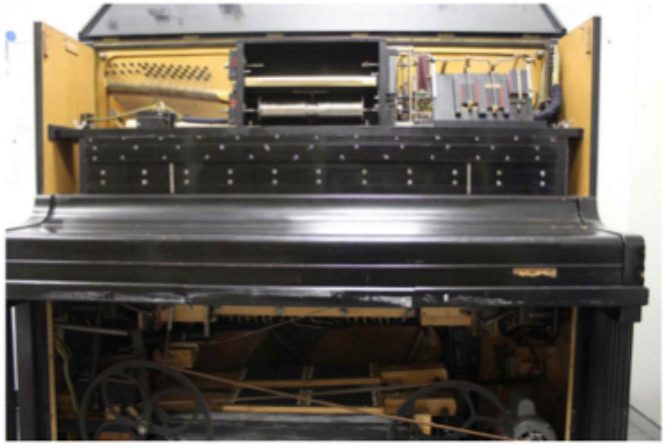
Symbolic Music Representations

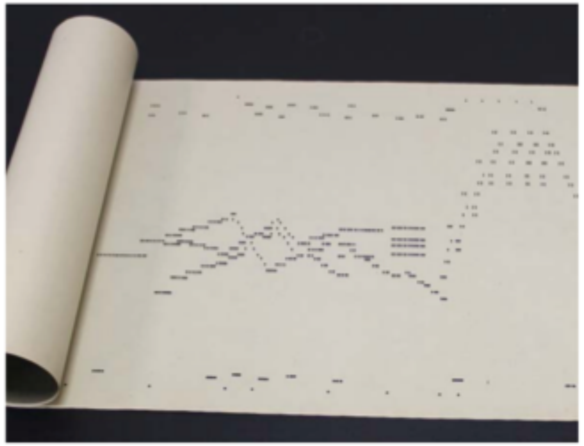
What is Symbolic Music Representation?

- Any score representation with explicit coding of notes or other musical events
- Includes:
  - Piano-roll representations
  - Midi Representations
  - Other Symbolic formats that encode sheet music

Piano Roll Representations

- Piano Roll has a long history from the late 19th century beginning with self playing pianos and is a continuous roll of paper with holes punched into it.
- The Holes represent notes and a note is triggered when a hole crosses the tracker bar
- Many famous pianists/composers like Gustav Mahler, Edvard Grieg, and George Gershwin have their recordings preserved on piano rolls



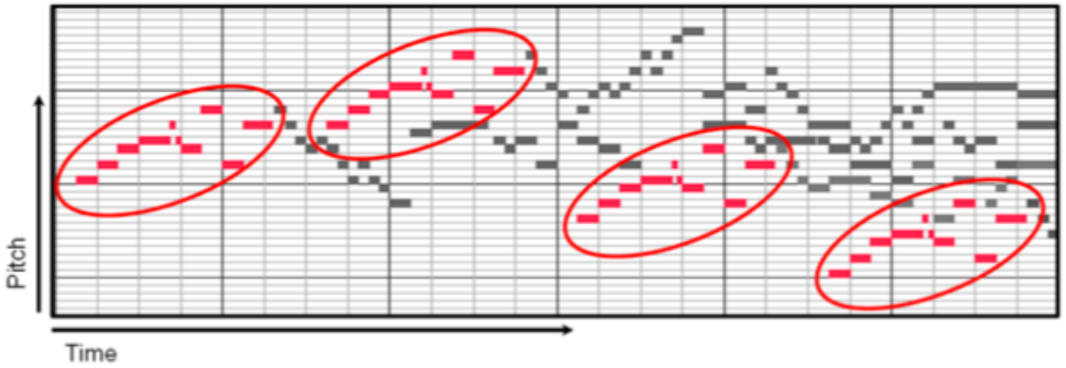


Piano Roll Representations

- Below is an example of a Bach Fugue BWV 846 in C Major encoded in piano roll vs the original sheet music
- Piano rolls are a big simplification of what is notated in sheet music but preserve important attributes of musical notes like pitch and durations

Sheet Music

Piano Roll



MIDI Representations

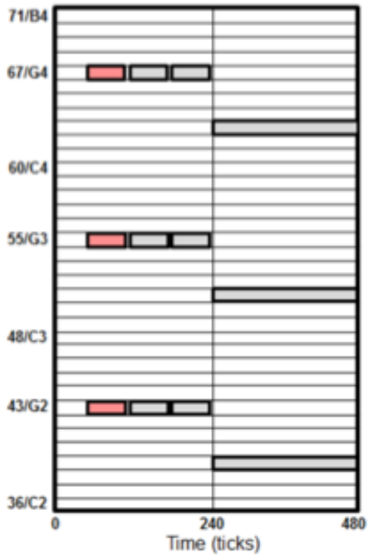
- MIDI – Musical Instrument Digital Interface
- Was originally used for getting electronic instruments to play based on MIDI signals
- Benefits of MIDI
  - Abundance of data online
  - Widespread usage over the last 3 decades

MIDI Representations

Note	Octave										
	-1	0	1	2	3	4	5	6	7	8	9
C	0	12	24	36	48	**60**	72	84	96	108	120
C#	1	13	25	37	49	**61**	73	85	97	109	121
D	2	14	26	38	50	**62**	74	86	98	110	122
D#	3	15	27	39	51	**63**	75	87	99	111	123
E	4	16	28	40	52	**64**	76	88	100	112	124
F	5	17	29	41	53	**65**	77	89	101	113	125
F#	6	18	30	42	54	**66**	78	90	102	114	126
G	7	19	31	43	55	**67**	79	91	103	115	127
G#	8	20	32	44	56	**68**	80	92	104	116	
A	9	21	33	45	57	**69**	81	93	105	117	
A#	10	22	34	46	58	**70**	82	94	106	118	
B	11	23	35	47	59	**71**	83	95	107	119	

- Typical Data Points in MIDI:
  - Time (How long to wait relative to previous message to play note)
  - Message (NOTE ON, NOTE OFF)
  - Channel (Different Instruments or Voicing)
  - Note Number (0-127 Controls Pitch)
  - Velocity (0-127 Controls intensity or volume of sound)
- Example from Beethoven of Midi Encoding:

Time (Ticks)	Message	Channel	Note Number	Velocity
60	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	67	100
0	NOTE ON	1	55	100
0	NOTE ON	2	43	100
55	NOTE OFF	1	67	0
0	NOTE OFF	1	55	0
0	NOTE OFF	2	43	0
5	NOTE ON	1	63	100
0	NOTE ON	2	51	100
0	NOTE ON	2	39	100
240	NOTE OFF	1	63	0
0	NOTE OFF	2	51	0
0	NOTE OFF	2	39	0

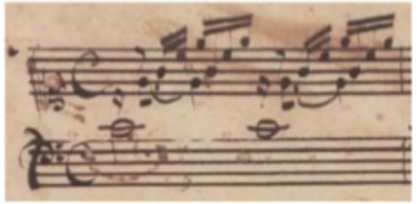


Score Representations

- “Score Representations yield explicit information about musical symbols such as the staff system, clefs, time signatures, notes, rests, accidentals, and dynamics"
- These representations are considered the closest to what is actually shown in sheet music in comparison to the previously mentioned MIDI files
- MusicXML – Typically the standard file used in Music notation applications by composers

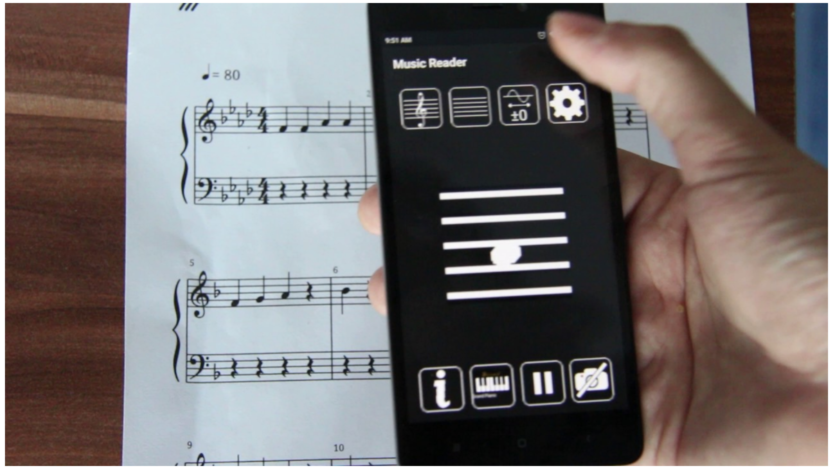
Computer Generated

Handwritten



Optical Music Recognition

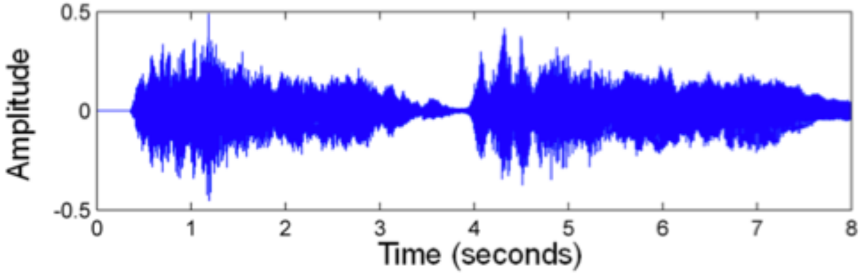
- The process of converting digital scans of printed sheet music into symbolic representations such as MIDI or MusicXML (Computer Vision)



Audio Representations

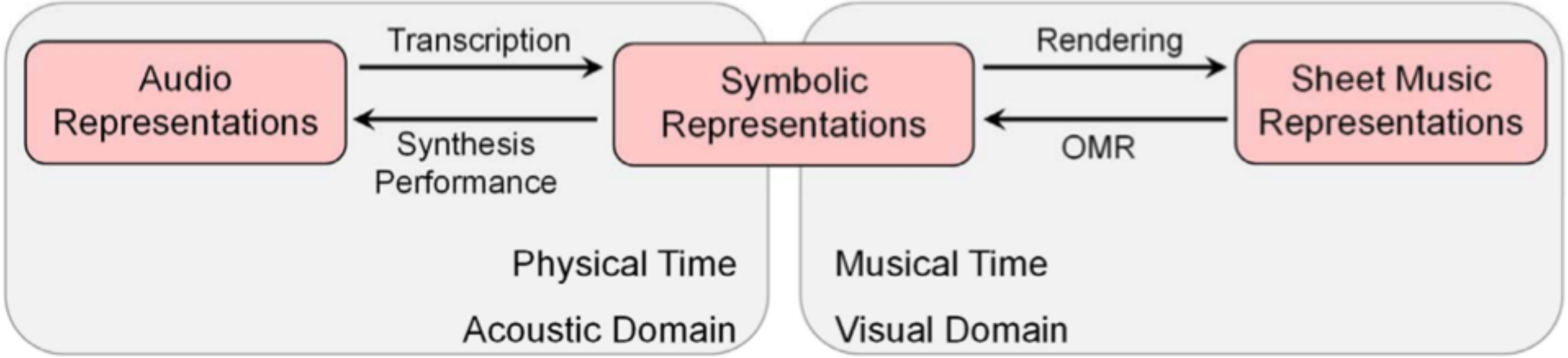
What is Audio Representation?

- Representation of acoustic sound waves
- Performing music results in sounds or acoustic waves which are traditionally converted to an electrical signal by microphone
- There are many elements of Audio:
  - Waves and Waveforms
  - Frequency and Pitch
  - Dynamics, Intensity, and Loudness
  - Timbre (tone color)



Summary

- There are three forms of Music Representations in Computational Music
- Symbolic Representations can often be seen as a link or bridge between Audio Representations (Sound) and Sheet Music Representations (Notation)



Deep Learning Techniques for Music Generation – A Survey

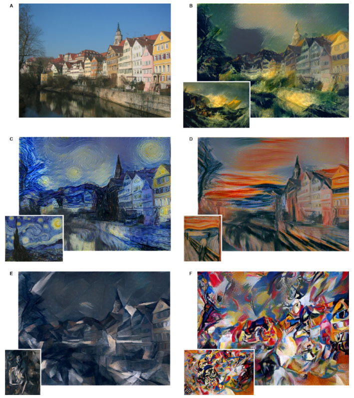
Jean-Pierre Briot, Gaetan Hadjeres, Fracois Pachet – Sep 2017

Preface

- “This book is a survey and an analysis of different ways of using deep learning (deep artificial neural networks) to generate musical content.”

Introduction

- This section discusses the recently popular Deep Learning techniques which are now routinely used for classification and prediction tasks such as image and voice recognition
- Success of Deep Learning attributed to:
  - Technical advances (pre-training/convolutions)
  - Availability of massive data
  - Dedicated computing power
- Deep Learning is usually applied to traditional Machine Learning tasks like classification and regression
- Recently it has been growing in the area of content generation
  - Images
  - Text
  - Music
  - Art



Book Roadmap

- **Chapter 2 - Method**
- **Chapter 3 - Objective**
- **Chapter 4 - Representation**
- Chapter 5 – Architecture
- Chapter 6 – Strategy
- Chapter 7 - Systems
- Chapter 8 – Analysis
- Chapter 9 – Other sources of inspiration
- Chapter 10 - Discussion

Chapter 2 - Method

- The Book defines 4 dimensions regarding deep learning for musical generation
  - Objective
  - Representation
  - Architecture
  - Strategy

1. Objective

- Type of Musical Content to be generated
  - Melody (Sequence of Notes)
  - Full Polyphony (Chorale) multiple voices
  - Accompaniment to given melody
  - The association of a melody and chords

2. Representation

- Representation – The type of data or information used to train or generate the musical content
  - Signal
  - Transformed Signal
  - Piano Roll
  - MIDI
  - Text
- Feature Extraction or handcrafted features may also be applied to improve generation

3. Architecture

- Architecture – Structure of the Neural Network
  - Neural Network Structure may be conventional with various layers or units
  - May Include convolutions
  - May be recurrent- to learn sequences
  - Autoencoder architectures are also useful for their ability to extract features
- Some systems combine architectures or architecture traits as seen later in this paper

4. Strategy

- Strategy – Method of using Deep learning Architectures to generate music
  - Direct use – Using a deep network architecture for prediction task (by feed forward computation) to produce music
  - Indirect Use - Most systems use this method
    - Sampling from a generated distribution
    - Input manipulation
    - Querying musical units from a generated description and concatenating them

Chapter 3 - Objective

- **Objective -** The type of musical content to be generated
  - **Melody** – Sequence of Notes
  - **Polyphony** – More than one voice or instrument
  - **Accompaniment** to a given melody
    - Counterpoint – (Melodies in Conjunction/concurrently)
    - Sequence of Chords – (Harmony)

- It is also important to consider the destination of generated music content
  - Computer to play the music content – (Audio / MIDI)
  - Human (Final Output must be a musical score)

Chapter 3 – Objective (Continued)

- Autonomy of the generation of musical content
  - Autonomous- Could be completely autonomous (automated- no human intervention)
  - Interactive Systems (Human user guiding the process of generation)
    - Interactive Systems are fairly new as Deep Learning Applied to music is also very new.
    - These function as support systems for musicians (composers, arranging, and other activities)
    - FlowComposer prototype
      - http://www.flow-machines.com/flowcomposer-composing-with-ai/

Chapter 4 - Representation

- Three Types/Stages of Data Representations
  - Training Input – Dataset used for training deep learning system
  - Generating Input – Data that is used for input generation – First Note / Melody for system to accompany
  - Generated Output – The output generated
- This paper then covers types of music data representations including – (Already Covered)
  - Sheet Music Representation
  - Symbolic Representation
  - Audio Representation
- Common Issues and Techniques with music representation
  - Challenges and how to pre-process Music Data for Neural Networks

Common Issues and Techniques with music representation

- **Global vs Time Slice**
  - “ The representation of time is fundamental for musical processes.”
  - Three cases mentioned:
    - Global – There is no such thing as temporal sequence/ explicit notion of time, Architecture is not recurrent, the granularity of the neural network processes the representation as a whole
    - Time step – the most frequent time representation, the granularity of training input or generation input is set to a temporal slice of the music content
    - Note step – The rarest case, the granularity of processing is a note

Common Issues and Techniques with music representation

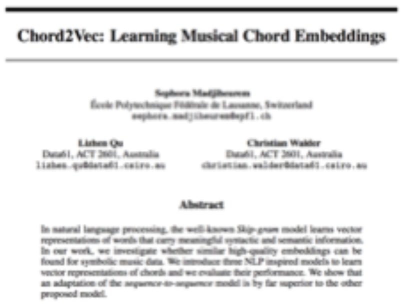
- **Note Ending**
  - One issue is the representation of the end of a note
  - In MIDI, Note Ending is represented by Note Off event
  - Piano Roll data does not have this representation
  - Strategies cited to address this include:
    - Always mark end of notes with special tags and divide by 2 the size of the time step
    - Special Computing Unit in the network to indicate beginning of a note

Common Issues and Techniques with music representation

- **Time Quantitization-** It is essential to define the value of the time step to allow the neural network to temporally interpret the representation.
- The Most common approach is to set the time step to the smallest time step in the entire dataset
  - If the smallest value is a 8^{th} note then the time step quantization is an 8^{th} note
  - This serves as the baseline measure of time

Common Issues and Techniques with music representation

- **Feature Extraction-** Preliminary step to generate features as an input for Neural Networks
- Representation of data in a more compact form – Gain Efficiency/Accuracy
- Handcrafted Features – Manually Created
  - Bag of Words (Generally applied to word data but also can be applied to music data)
- Autogenerated features – via Autoencoder
  - Word2Vec model – a recent model for natural language, Chord2vec is a similar model of vector encoding of chords.



Chord2Vec: Learning Musical Chord Embeddings

Sephora Madjiheurem
École Polytechnique Fédérale de Lausanne, Switzerland
sephora.madjiheurem@epfl.ch

Lizhen Qu
Data61, ACT 2601, Australia
lizhen.qu@data61.csiro.au

Christian Walder
Data61, ACT 2601, Australia
christian.walder@data61.csiro.au

Abstract

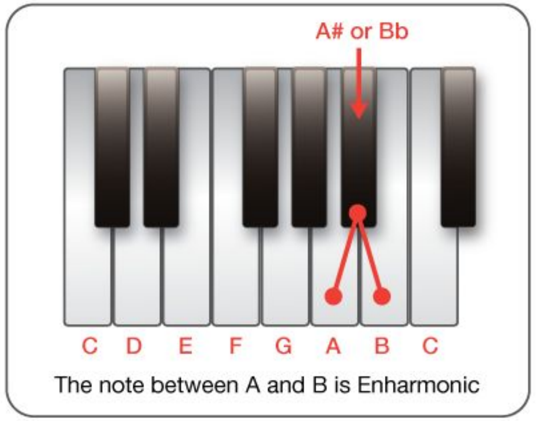
In natural language processing, the well-known Skip-gram model learns vector representations of words that carry meaningful syntactic and semantic information. In our work, we investigate whether similar high-quality embeddings can be found for symbolic music data. We introduce three NLP inspired models to learn vector representations of chords and we evaluate their performance. We show that an adaptation of the sequence-to-sequence model is by far superior to the other proposed model.

Common Issues and Techniques with music representation

- **Input Encoding**
  - This challenge is regarding how to represent a set of variables like pitch into a set of inputs for the neural network
- Value encoding **-** Encode a numerical value for input node
  - Pitch – Hertz Frequency Value, Corresponding MIDI pitch, Relative value within the interval (Normalization)
- One hot encoding – Each Possible Note is considered a distinct element of a vocabulary
  - N Input Nodes – N is size of vocabulary (Number of distinct notes)
  - Presence of a note, not the note value will be encoded
  - Presence of Pitch will be encoded as 1 and others will be encoded as 0 – Sparsity Matrix
- One-Hot Encoding is the most frequently used strategy
  - Reformulates Prediction Task of a note value to Classification task between set of values

Common Issues and Techniques with music representation

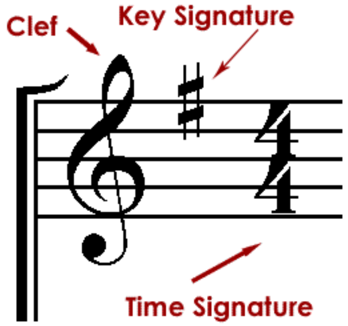
- **Note Encoding**
  - Most systems consider enharmony, Pitches like A# and Bb to be equivalent
  - MIDI data also represents these as the same number
  - An exception is BachBot which encodes notes using their real name
    - DeepBach states this leads to more accurate model and better result
    - http://bachbot.com/



© 2012 www.scaletrainer.com

Common Issues and Techniques with music representation

- **Meta-Data**
  - Additional data from scores may be represented
    - Key, time signature, note ties, fermata, harmonics. Etc.
    - This extra information helps with learning and generation



Common Issues and Techniques with music representation

- **Transposition**
  - A common technique in Machine Learning is generating synthetic data to augment the size of the dataset
  - In the musical domain, this can be easily done by transposing each training example into every key
    - This reduces data sparsity and makes examples more generic
    - Think of Karaoke (Can change the key of song but musical context stays intact)
  - This approach has been utilized in a few papers along with the opposite approach of transposing all training samples to one key

Common Issues and Techniques with music representation

- **Datasets**
  - Unfortunately there are not reference datasets like MNIST for the music domain
  - The authors do specify some datasets that other papers have used to generate music:
    - Polyphonic Music – JSB Chorales dataset
    - Symbolic Music – Music Data by Walder, MusicNet,
    - Lead Sheets – LSDB (Lead Sheet Data Base)

Next Presentation- Architecture

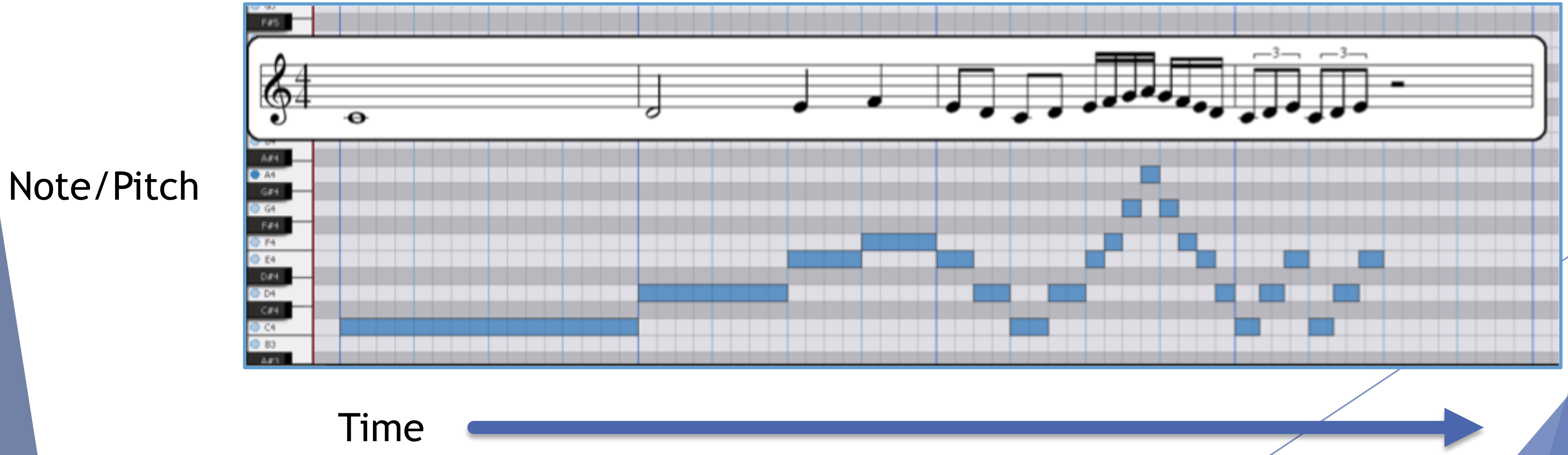
- This section discusses many architectures or neural network structures being applied to deep learning for Music
  - Recurrent Neural Network (RNN)
  - Long Short-Term Memory (LSTM)
  - Autoencoder
  - Stacked Autoencoders
  - Restricted Boltzmann Machine (RBM)
  - Variational Autoencoder
  - Convolutional Architectural Pattern
  - Conditioning Architectural Pattern
  - Generative Adversarial Networks (GAN)
  - Reinforcement Learning
  - Compound Architectures
    - Convolutional Generative Adversarial Networks
    - Recurrent Generative Adversarial Networks
    - Others

Thesis

My Interest

- My interest largely lies in Symbolic Music representation
  - MIDI files
  - Piano Roll
- MIDI files are often fundamental data files for composers/film scorers nowadays

Piano Roll vs Traditional Notation



Thesis Idea

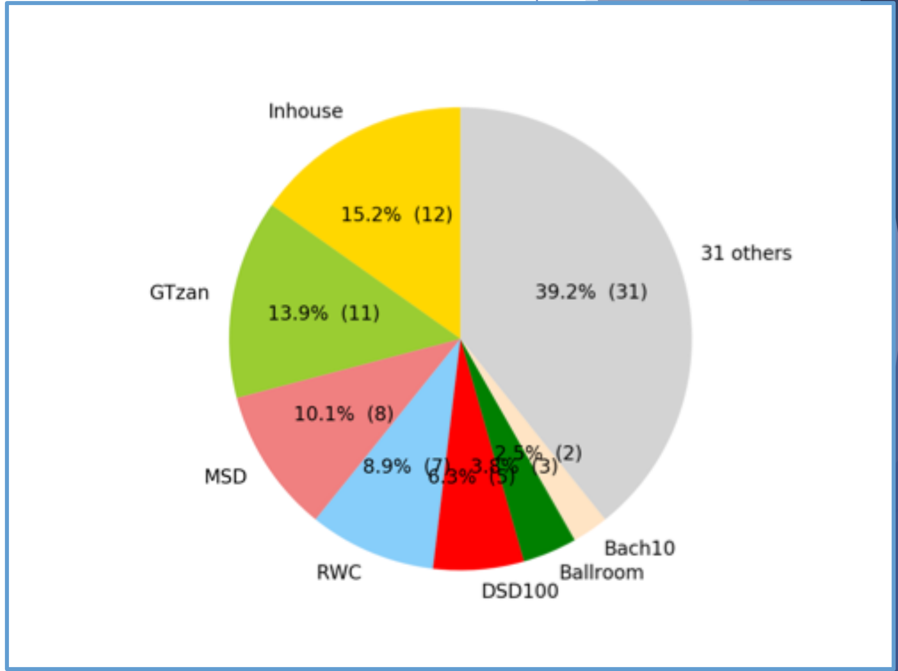
- **I am still in the process of researching an Evaluation Method**
  - A few papers used Survey and 5 Point Likert Scale to evaluate musical outputs.
- **Method 1** – Create multiple models based on recently used music generation algorithms and use the same training set (Markov Chain, Restricted Boltzmann Machines, RNN/LSTM)
  - Evaluate Musical Output via survey as music is subjective
  - Determine which model is best in eyes of listeners
- **Method 2** – Create one model based on RNN/LSTM based on one composers data.
  - Evaluate Musical Output via survey as music is subjective
  - Determine whether this music is plausible/pleasant

Steps

- 1. Research History of Algorithmic Music Composition/Deep Learning
- 2. Gather Data Files (Midi Files/Symbolic Representations of Music)
- 3. Clean Data (Format Data and feature engineer so data is proper for NN)
- 4. Train the Model/Perform Experiments (Using NN's etc.)
- 5. Evaluate Model performance

Data Sources

- Touhou:
  - http://easypianoscore.jp/
- Final Fantasy:
  - https://www.thefinalfantasy.com/site/midi-collection.html
- Bach/Chopin
  - https://www.classicalarchives.com/midi.html



Most Common Datasets used in DL Music Research

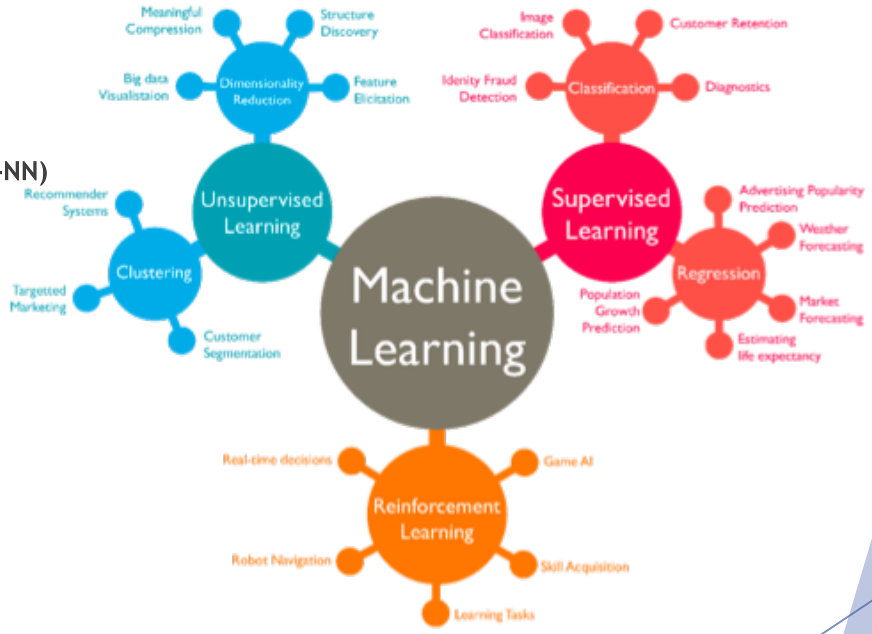
Literature Review

After Performing Research on Deep Learning for Music Generation, I found many resources that I will review in the next 2 years.

- https://medium.com/artists-and-machine-intelligence/neural-nets-for-generating-music-f46dffac21c0 - Neural Nets application to music
- http://www.hexahedria.com/2015/08/03/composing-music-with-recurrent-neural-networks/ - Composing with RNN
- http://people.idsia.ch/~juergen/blues/IDSIA-07-02.pdf - Composing with LSTM
- https://cs224d.stanford.edu/reports/allenh.pdf - Recent paper on Deep Learning for Music
- https://magenta.tensorflow.org/ - **Magenta** is a research project exploring the role of machine learning in the process of creating art and music.
- https://arxiv.org/pdf/1709.01620.pdf - Deep Learning for Music Generation Survey Paper
- https://github.com/ybayle/awesome-deep-learning-music - Compilation of Deep Learning related Music research

Research Plan

- Year 1: (General Machine Learning)
  - 11/9 - Machine Learning - Decision Tree’s / Random Forest
  - 12/7 - Machine Learning Basics / End to End ML / Regression
  - 1/18 - Machine Learning - Classification (Logistic Regression, LDA, K-NN)
  - 4/25 - Machine Learning - Recommender Systems
  - Machine Learning – Naïve Bayes Classifier/ Bayes Theorem
  - Machine Learning – Support Vector Machines
  - Machine Learning - Natural Language Processing
  - Machine Learning - Clustering
  - Machine Learning - Deep Learning
  - Machine Learning – XGBoost (New)
  - Machine Learning – Feature Engineering (New)
- Thesis – Deep Learning for Music Generation
  - Music Composition (Deep Learning)



Github

- https://steimel64.github.io/