Drum fills detection and generation

Frédéric Tamagnan & Yi-Hsuan Yang

Introduction

Origins

We wanted to generate drum fills as an answer to regular patterns with Deep Learning

↓

We needed data

↓

We had to detect and extract drum fills

What is a drum fill ?

https://www.youtube.com/embed/u5MIa4wgmU4?start=140&enablejsapi=1

Why to detect and generate drum fills ?

1. To segment a music piece
2. To make long-term music generation with dynamic and variations
3. To make short-term music generation for live performances

Why to detect and generate drum fills ?

Kink, boiler room Moscow, Live set, 2015

Why to detect and generate drum fills ?

Tr-8S, Roland

Why to detect and generate drum fills ?

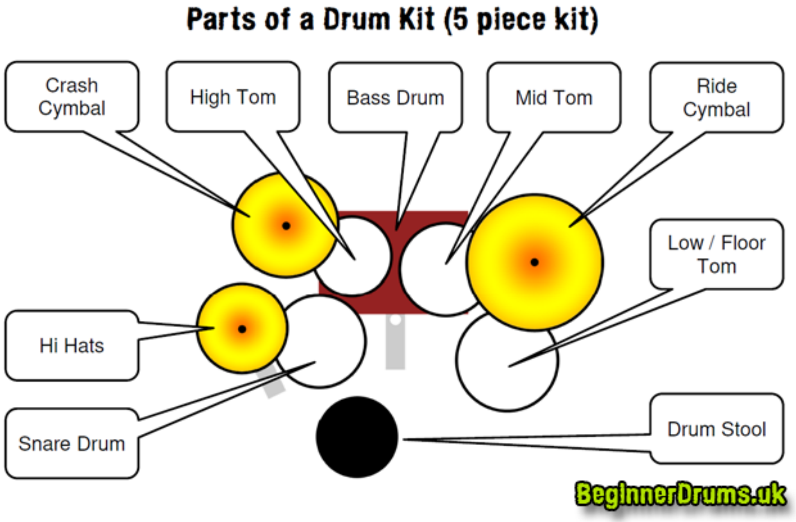
Tr-8S, Roland

Challenges

- Hard to **define** what is a drum fill with a **general rule**
- **No big datasets** with drum fills labels

Problem definition

- We focus on **detection and generation of 4/4 bars containing a drum fill**
- We **don't take in account** the **precise boundaries** of the fills
- We use **9 instruments * 16 timesteps tensor** to represent a drum bar



Empirical observations

Drum fills :

1. A **greater use of toms, snares or cymbals**, than in the regular drums pattern
2. A **difference of played notes** between the regular pattern and the drum fill
3. An appearance in general **at the end of a cycle of 4 or 8 bars**

Datasets at our disposal

1. **Labelled dataset** : Native instruments + Oddgrooves.com midi drums pack : **5,317 regular patterns bars + 1,1412 drum fills bars**

2. **Unlabelled dataset** : Lakh pianoroll dataset : **21,425 songs** with their related pianorolls

Dong, H.W.,Hsiao, W.Y., Yang, L.C., Yang, Y.H.: MuseGAN: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In:Thirty-Second AAAI Conference on Artificial Intelligence (2018)

Drum fills Detection

Drum fills Detection

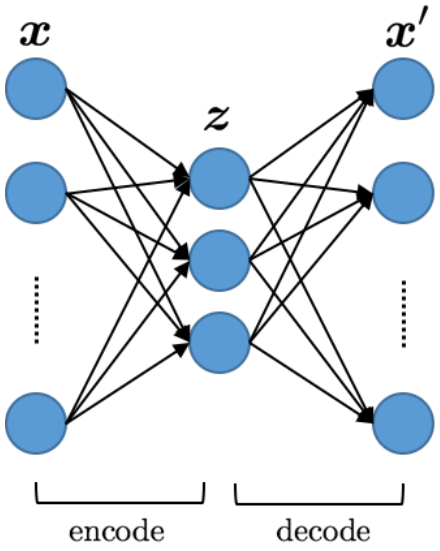
2 Methods

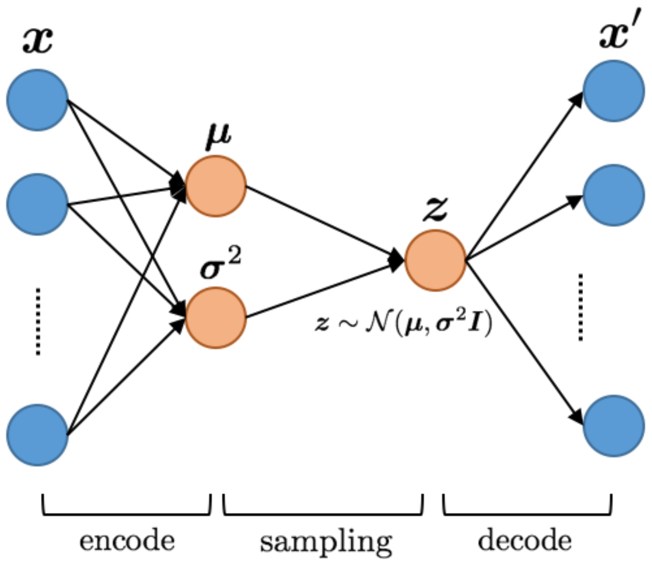
- Supervised Learning
- Rule-based Method

Supervised Learning

Features

- Variational Auto-encoder latent space features

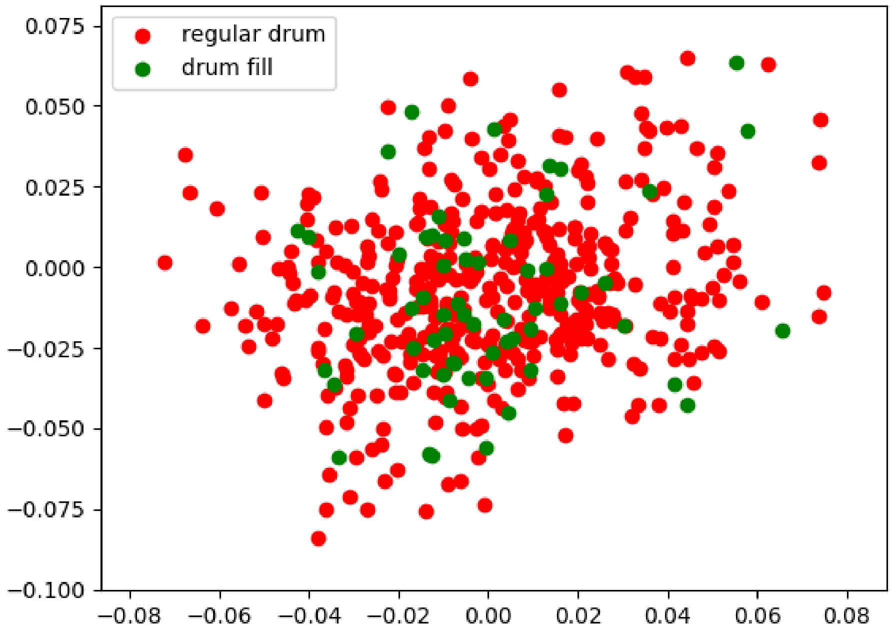




Supervised Learning

t-SNE Visualization

Drum fills and regular patterns in the **latent space of a VAE trained on the LDP dataset**

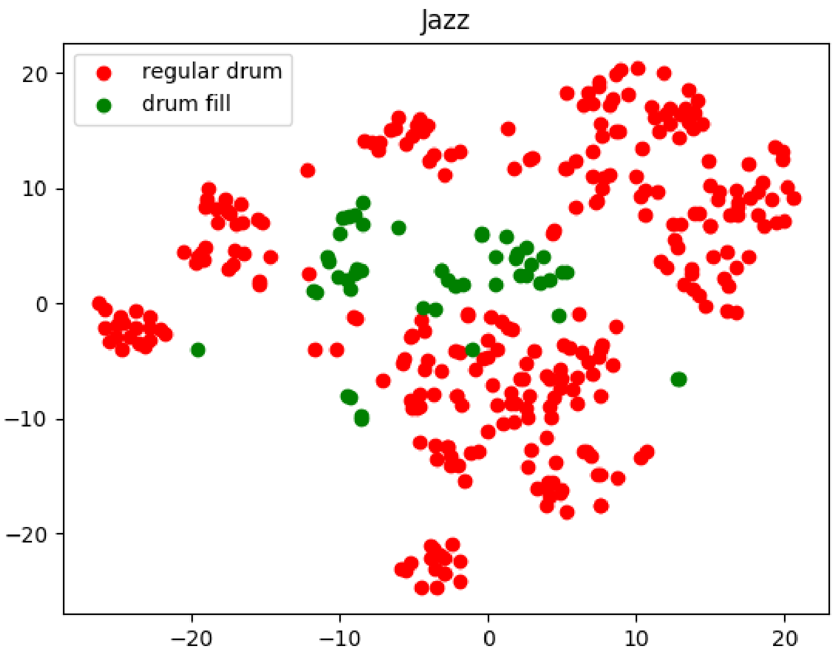


Hard to separate if we consider all the bars at the same time !

Supervised Learning

t-SNE Visualization

Drum fills and regular patterns in the **latent space of a VAE trained on the LDP dataset**

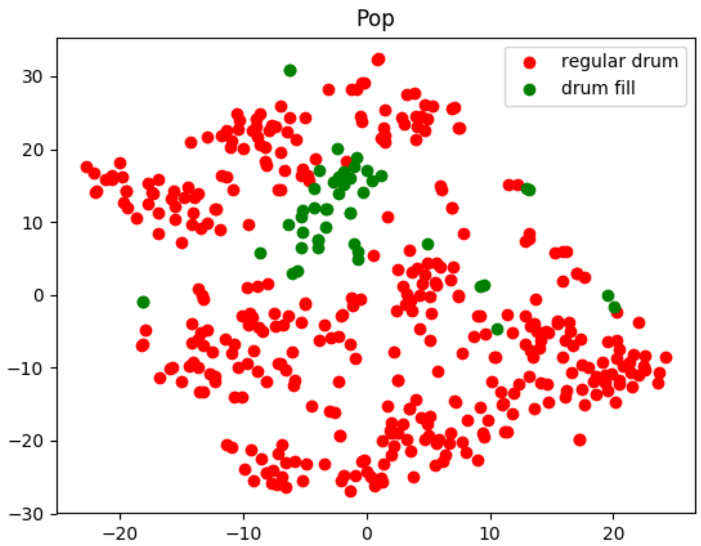


Better if we consider only one genre !

Supervised Learning

t-SNE Visualization

Drum fills and regular patterns in the **latent space of a VAE trained on the LDP dataset**

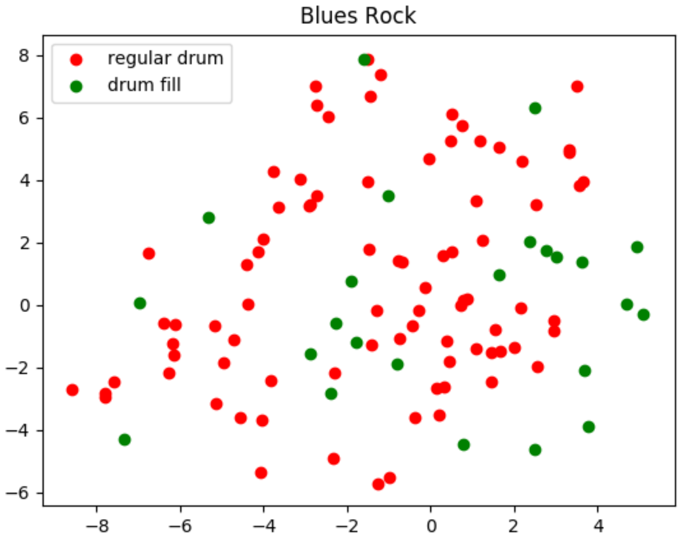


Better if we consider only one genre !

Supervised Learning

t-SNE Visualization

Drum fills and regular patterns in the **latent space of a VAE trained on the LDP dataset**



...but not always the case

Supervised learning

features

- VAE latent space features

  +

  Handcrafted Features :
- Instruments used
- Max, std, mean of velocity

  = Dimension of input vector : 59

Supervised Learning

Model

- Logistic Regression
- Standardization
- L2 Regularization

Supervised Learning

Validation

Feature set	Precision	Recall	F1 Score
HD	0.80	0.79	0.79
LS	0.58	0.06	0.10
HD+LS	0.89	0.81	0.85

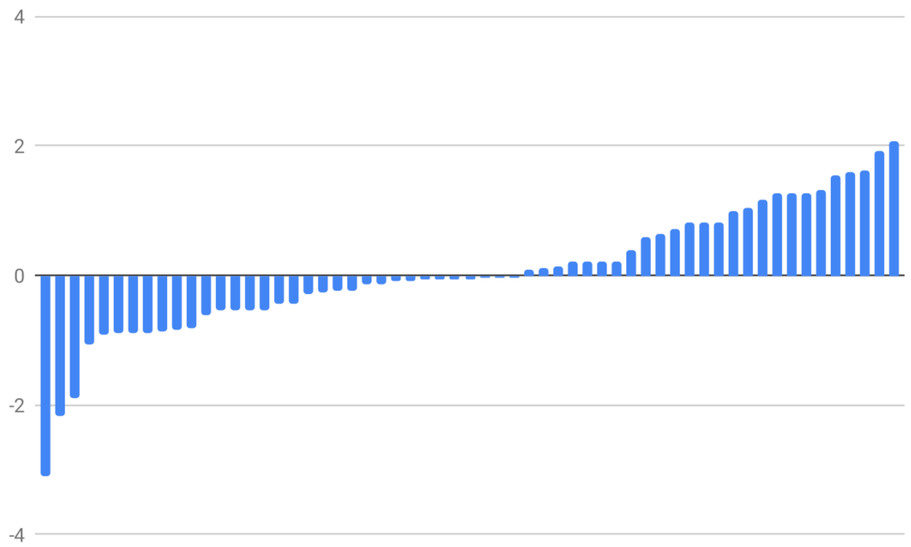
Table 2. Validation metrics of our classifier. HD : Handcrafted features, LS : VAE's latent space features

NB : Handcrafted features : Velocity features + use of instruments

Supervised Learning
Validation

Most correlated Hand-crafted features :

1. max velocity of high tom,
2. Std of velocity of mid tom
3. max velocity of low tom



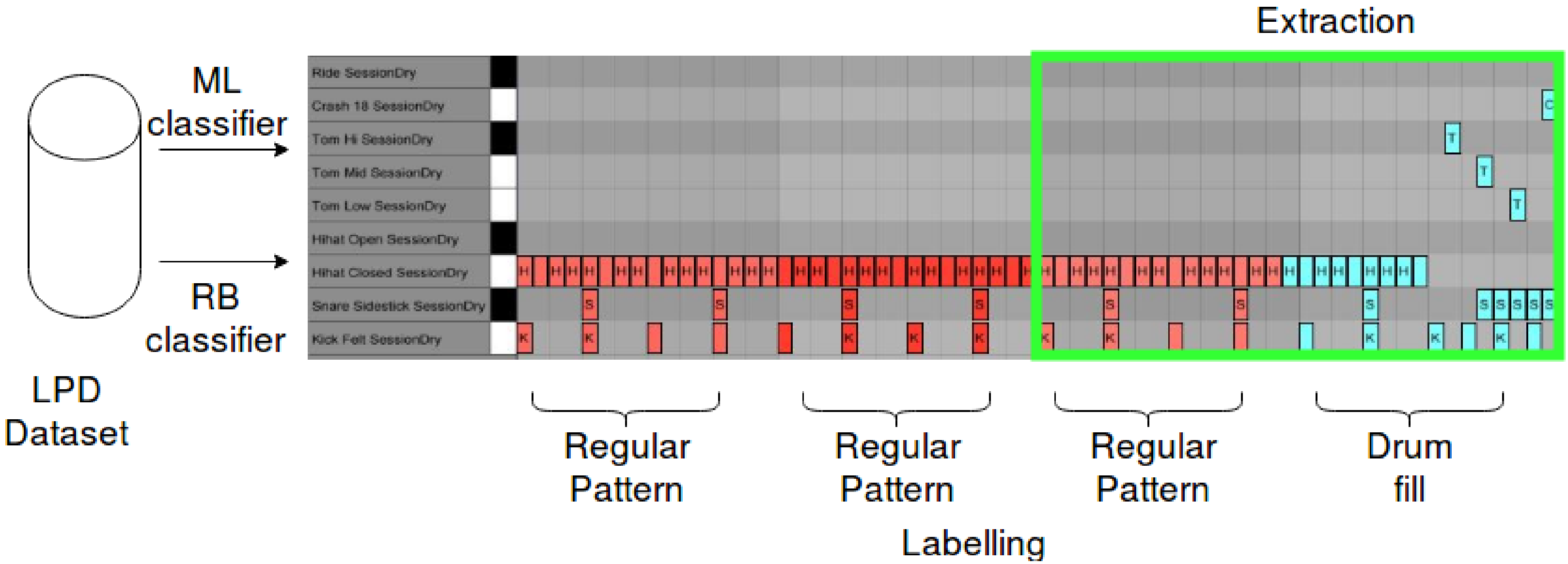
Rule-based Method

Difference of notes between two bars

Let A, B two bars binarized tensors of dimensions $t \times n$ (time steps $\times$ number of instruments), we define the difference of notes DN between A and B as:

$$DN(A, B) = \sum_{\substack{0 \leq i < t \\ 0 \leq j < n}} max(0, A_{i,j} - B_{i,j}) \qquad (1)$$

Labelling and extraction

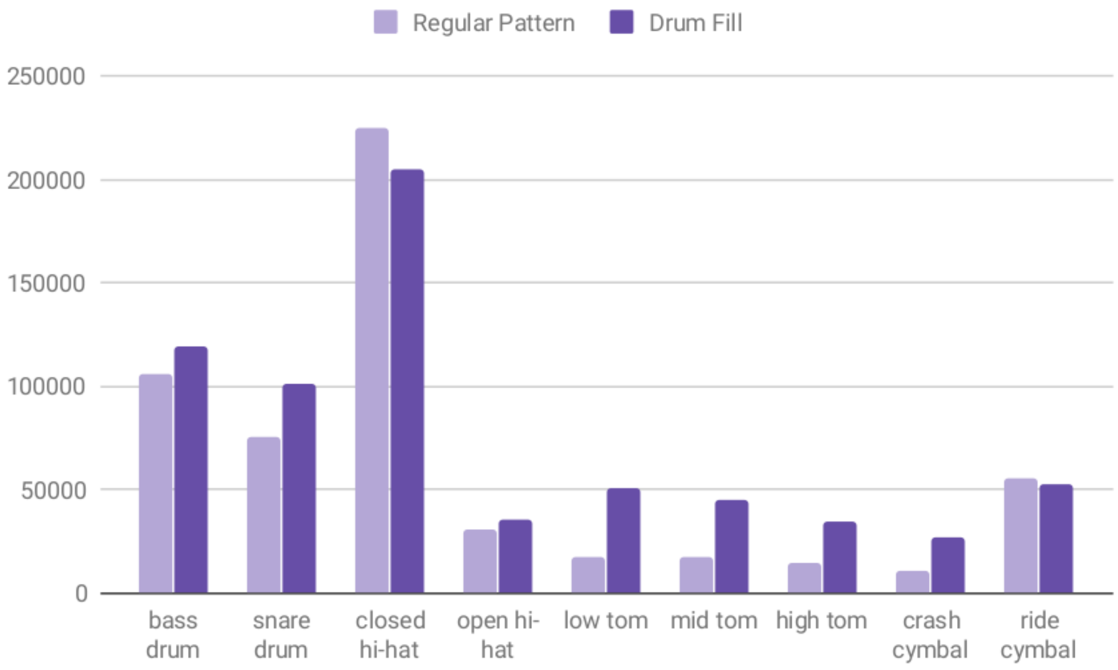


Data cleaning

- Removing duplicated rows
- Removing all the couples where the regular pattern or the drum fill have fewer than 7 notes
- Removing all the couple where the drum fill has a too high density of snare notes, above 8

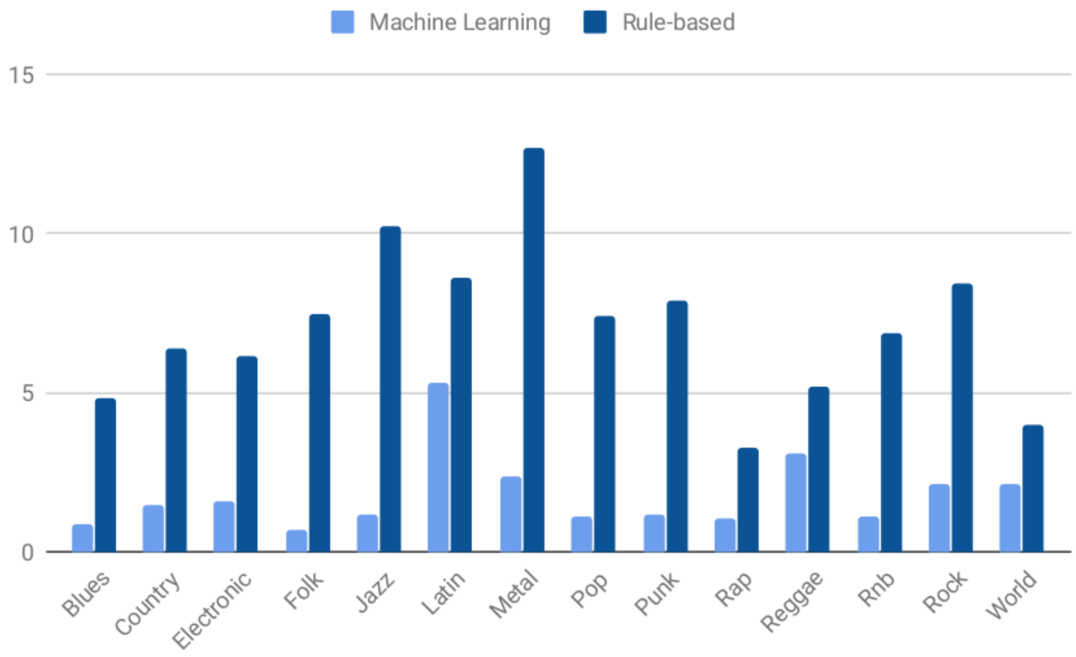
	#ML dataset	**#RB dataset**
Raw	13,476	97,023
After rule 1	6,324	45,723
After rule 2	5,271	39,108
After rule 3	3,283	32,130

Extraction Evaluation



Amount of notes by instrument for the ML dataset

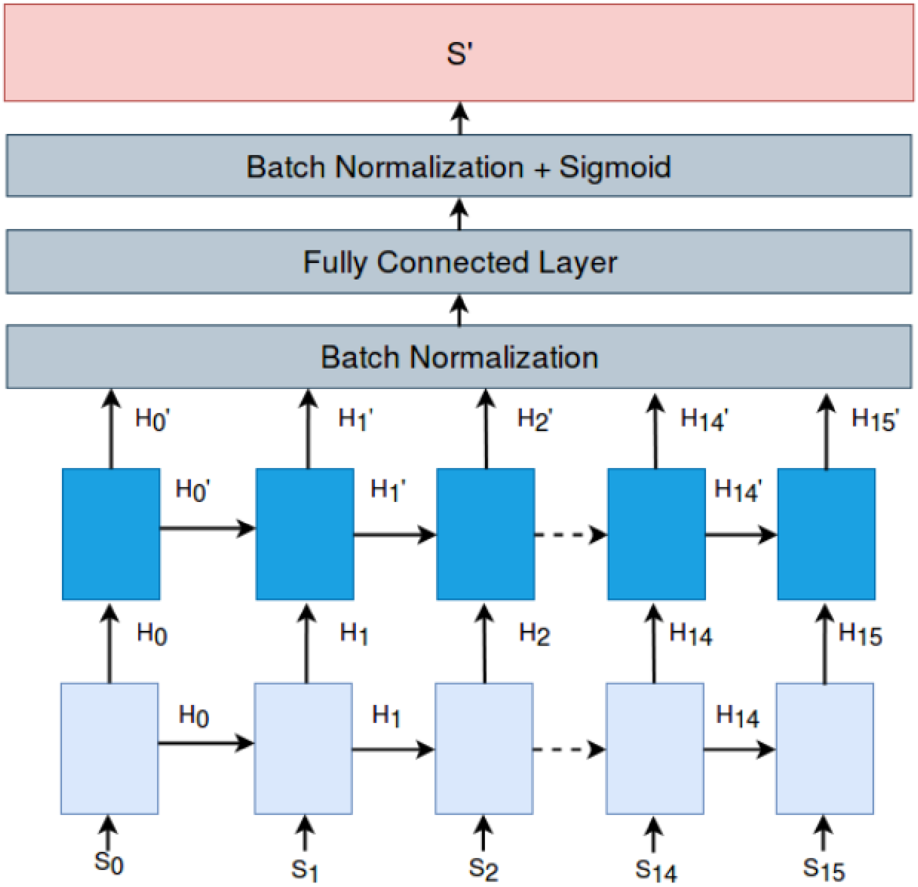
Extraction Evaluation



Drum fills Generation

Generation

RNN Many-to-many



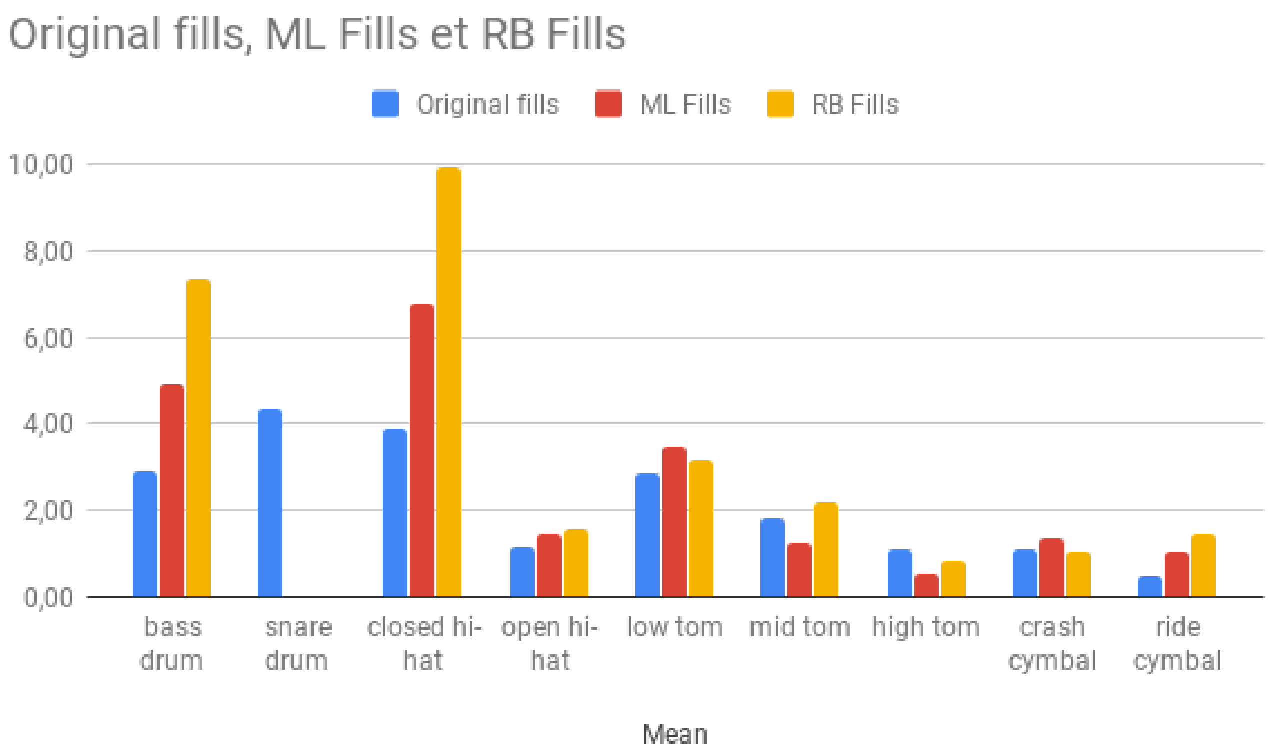
Input : Regular pattern bar

Output : Drum fill bar

Generation

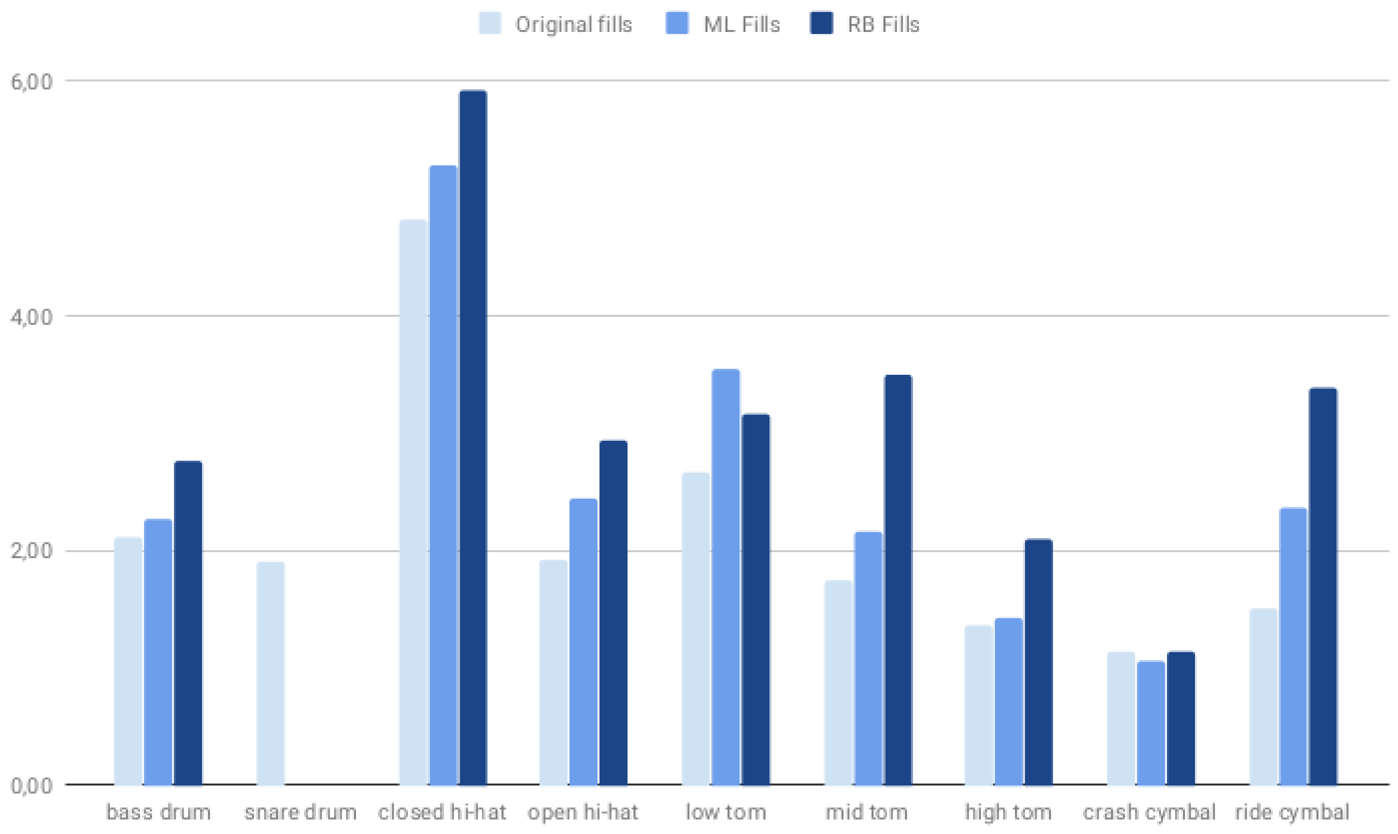
Evaluation

Mean of notes by instrument



Generation

Evaluation
Standard-deviation by instrument



Generation

Evaluation
Euclidian distance in the latent spae

	Sum of euclidean distance
ML fills	93012
RB fills	93844
Original fills	102135

Génération

Evaluation
User Study

- 51 participants
- 50% amateur musicians
- 14% semi-professional musicians
- 2% professional musicians

Among musicians :

- 78% DAW users
- 53% drummers

Génération

Evaluation
User Study

We asked people to compare :

- 1 ML fill
- 1 RB fill
- 1 Original fill (ground truth)
- 1 Rule composed fill (same layer of cymbals and toms applied on the regular pattern)

Generation

Evaluation
User Study

https://w.soundcloud.com/player/?url=https%3A//api.soundcloud.com/playlists/797390628&color=%23ff5500&auto_play=false&hide_related=false&show_comments=true&show_user=true&show_reposts=false&show_teaser=true

Generation

Evaluation

User Study

	ML	RB	Original	RC
Overall grade	2.61	2.90	**3.13**	3.10
Most coherent	17%	18%	29%	**36%**
Less coherent	**30%**	**30%**	23%	18%
Best groove	13%	25%	**34%**	28%
Worst groove	**35%**	30%	18%	17%

Generation

Evaluation
User Study

Why the results are bad, even for the human fills ?

- Hard to evaluate a fill with no musical background playing
- Specific and complex notion
- Only five sets of examples
- Hard to give a rating about a really short event
- …

Future directions

- Train a classifier with handlabelled data
- Use of binary neurons
- More sophisticated generation method

Thank you for your attention !

Mail : frederic.tamagnan@gmail.com