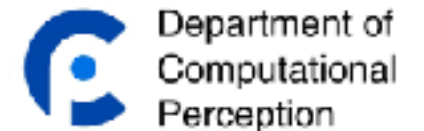


# FROM DRUM TRANSCRIPTION TO DRUM PATTERN VARIATION

Richard Vogl  
richard.vogl@tuwien.ac.at



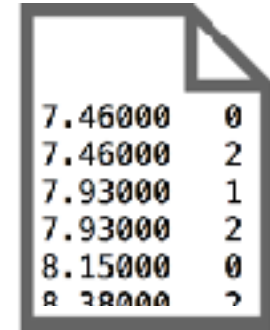
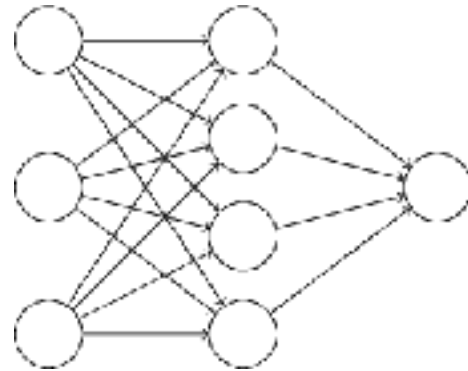
mir group



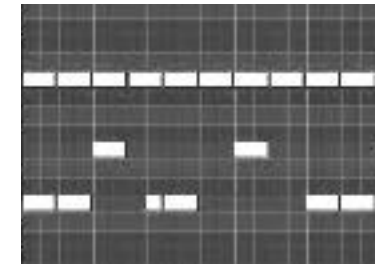
# PART 1

# AUTOMATIC DRUM TRANSCRIPTION

# WHAT IS DRUM TRANSCRIPTION?



7.46000	0
7.46000	2
7.93000	1
7.93000	2
8.15000	0
8.38000	2



- **Input:** popular music containing drums
- **Output:** symbolic representation of notes played by drum instruments

# STATE OF THE ART

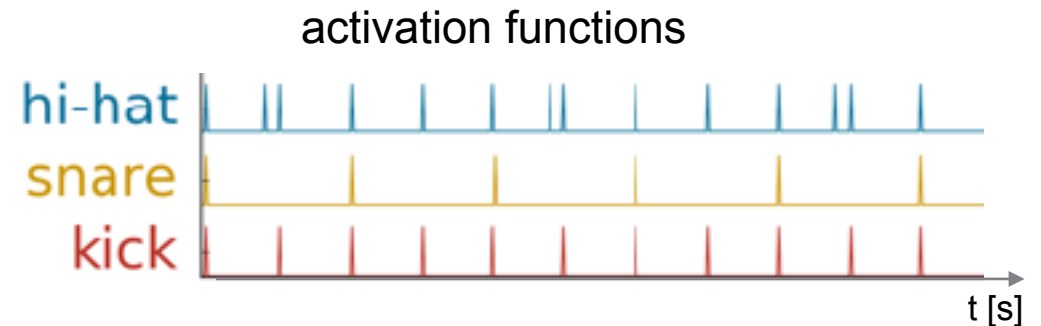
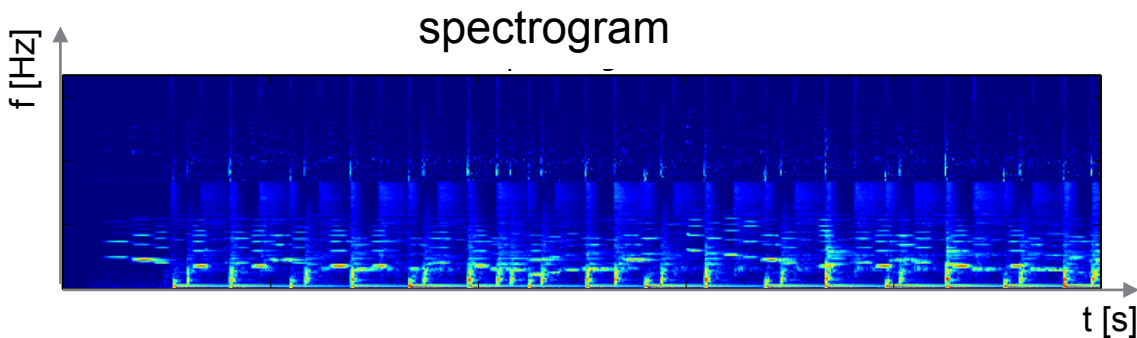
## ■ Overview Article

*Wu, C.-W., Dittmar, C., Southall, C., Vogl, R., Widmer, G., Hockman, J., Müller, M., Lerch, A.:*

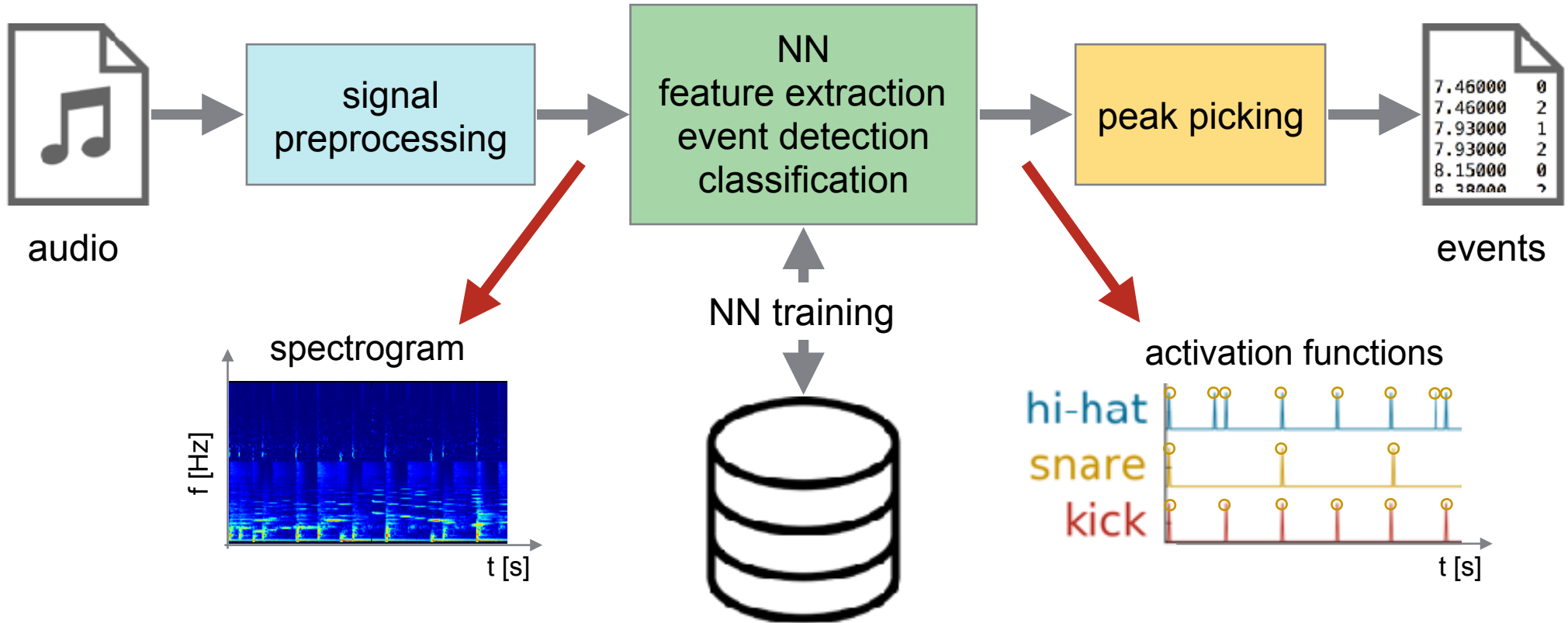
**“An Overview of Automatic Drum Transcription,”** IEEE Trans. on Audio, Speech and Language Processing, vol. 26, no. 9, Sept. 2018.

## ■ Current state-of-the-art systems:

- ▶ End-to-end / **activation-function-based** approaches
- ▶ **NMF** based approaches and **NN** approaches



# SYSTEM OVERVIEW



# PUBLIC DATASETS

## ■ IDMT-SMT-Drums [Dittmar and Gärtner 2014]

- ▶ Solo drum tracks, recorded, synthesized, and sampled
- ▶ 95 tracks, total: **24m**, onsets: **8004** + training samples



SMT solo

## ■ ENST-Drums [Gillet and Richard 2006]

- ▶ Recordings, three drummers on different drum kits, **optional accompaniment**
- ▶ 64 tracks, total: **1h**, onsets: **22391** + training samples



ENST solo



ENST acc.

# PERFORMANCE

- Simple RNNs architecture (GRUs)
- With **data augmentation**
- New state-of-the-art on public datasets (ICASSP'17):

F-measure [%] for individual methods on datasets

Method	SMT solo		ENST solo	ENST acc.
NMF	—	(95.0)	—	—
PFNMF	81.6	(—)	77.9	72.2
HMM	—	(—)	81.5	74.7
BDRNN	83.3	(96.1)	73.2	66.9
<b>tsRNN</b>	<b>92.5</b>	<b>(96.6)</b>	<b>83.3</b>	<b>75.0</b>

Richard Vogl, Matthias Dorfer, and Peter Knees, “**Drum transcription from polyphonic music with recurrent neural networks**,” in *Proc. 42nd IEEE Intl. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, LA, USA, Mar. 2017.

# ISSUES OF CURRENT SYSTEMS

- Performance not satisfying on real music
- Do not produce additional information for transcripts  
***drum onset detection*** vs ***drum transcription***

- ▶ bars lines

- ▶ tempo

- ▶ meter

- ▶ dynamics / accents

- ▶ stroke / playing technique

- Limited to three instrument classes

- etc.

ROCK - STRAIGHT 8THS ♩ = 192

(2+2-2+2+3+3)

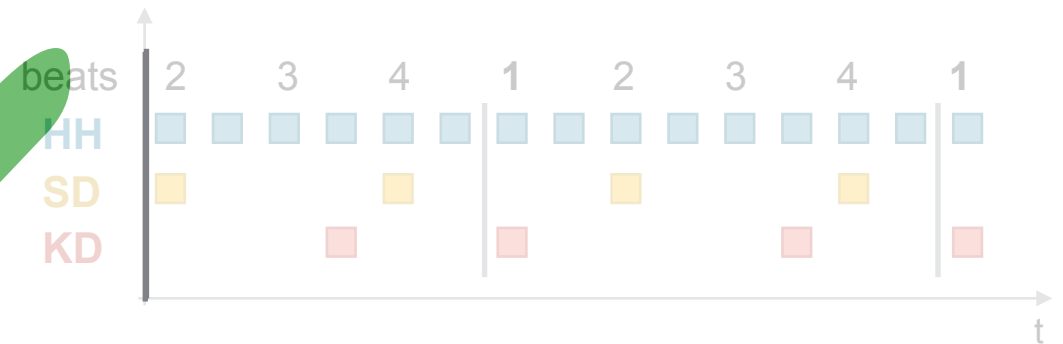
8 CLOSED HAT



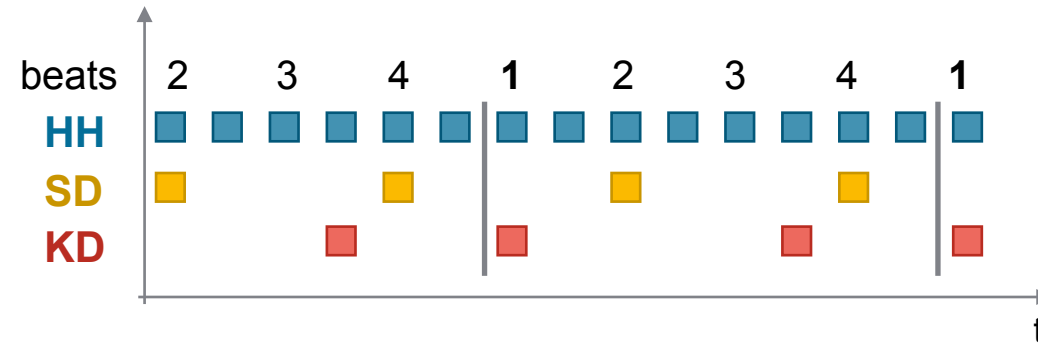
# ADDITIONAL INFORMATION FOR TRANSCRIPTS

■ Use beat and downbeat tracking to get:

- ▶ bars lines
- ▶ tempo
- ▶ meter



# LEVERAGE BEAT INFORMATION



- Beats are **highly correlated** with drum patterns
- Assume that **prior knowledge** of beats is helpful for drum transcription (drum hit locations / repetitive patterns)
- Use **multi-task learning** for beats and drums

# NEW DATASETS (DRUMS AND BEATS)

NEW!

**RBMA13-Drums** [<http://ifs.tuwien.ac.at/~vogl/datasets/>]

- ▶ Free music from the 2013 Red Bull Music Academy, different styles
- ▶ 27 tracks, total: **1h 43m**, onsets: **24365**
- ▶ **drum, beat, and downbeat** annotations



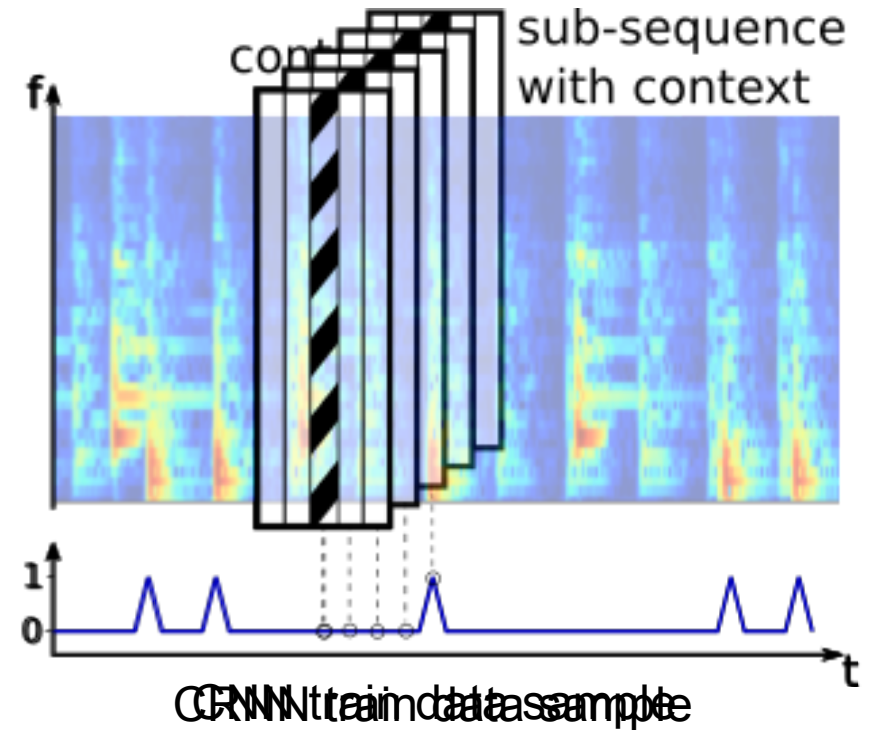
# CONVOLUTIONAL RECURRENT NN MODELS

## ■ Convolutional NN (CNN)

- ▶ Convolutions capture **local correlations**
- ▶ **Acoustic modeling** of drum sounds

## ■ Convolutional RNN (CRNN)

- ▶ **"best of both worlds"**
- ▶ Low-level CNN for **acoustic modeling**
- ▶ Higher-level RNN for **repetitive pattern modeling**



# PERFORMANCE

- New state-of-the-art using CRNNs (ISMIR'17)
- Multi-task learning can improve performance (for recurrent architectures):

	SMT	ENST		RBMA13			
		solo	acc.	DT	BF	MT	
<i>GRUs</i> [36]	92.5	83.3	75.0	-	-	-	
BGRU-a	93.0	80.9	70.1	59.8	63.6	64.6	RNNs
BGRU-b	93.3	82.9	72.3	61.8	64.5	64.3	
CNN-a	87.6	78.6	70.8	66.2	66.7	63.3	CNNs
CNN-b	93.4	<b>85.0</b>	78.3	66.8	65.2	64.8	
CBGRU-a	<b>95.2</b>	84.6	76.4	65.2	66.1	66.9	CRNNs
CBGRU-b	93.8	83.9	<b>78.4</b>	<b>67.3</b>	<b>68.4</b>	<b>67.2</b>	

Richard Vogl, Matthias Dorfer, and Peter Knees, “**Recurrent neural networks for drum transcription**,” in *Proc. 17th Intl. Soc. for Music Information Retrieval Conf. (ISMIR)*, New York, NY, USA, Aug. 2016.

# MIREX'17 RESULTS

Algorithm	mean fm	mean pr	mean rc	BD mean fm	SD mean fm	HH mean fm	
CW1	0.51	0.46	0.68	0.68	0.48	0.38	} NMF
CW3	0.53	0.50	0.65	0.67	0.46	0.42	
CW2	0.55	0.52	0.66	0.70	0.55	0.40	
RV3	0.68	0.74	0.70	0.81	0.64	0.51	RNN
RV2	0.67	0.69	0.73	0.78	0.67	0.51	CNN
RV1	0.71	0.75	0.74	0.82	0.70	0.53	CRNN
RV4	0.70	0.74	0.73	0.81	0.70	0.52	ensemble
CS1	0.61	0.56	0.73	0.79	0.55	0.46	} RNN
CS3	0.63	0.59	0.75	0.78	0.58	0.49	
CS2	0.63	0.61	0.71	0.78	0.57	0.49	

[http://www.music-ir.org/mirex/wiki/2017:Drum\\_Transcription\\_Results](http://www.music-ir.org/mirex/wiki/2017:Drum_Transcription_Results)

# EXAMPLES

RBMA13 Track 18



Original



Drums



Mixed

RBMA13 Track 15



Original



Drums



Mixed

# MORE DRUM INSTRUMENTS!

■ More complete and detailed transcripts

■ Challenges

- ▶ Not well defined / context dependent
- ▶ Similar sounds
- ▶ Diversity of sounds of certain instruments

number of classes			instrument name
3	8	18	
BD	BD	BD	bass drum
SD	SD	SD	snare drum
		SS	side stick
		CLP	hand clap
		HT	high tom
	TT	MT	mid tom
		LT	low tom
		CHH	closed hi-hat
HH	HH	PHH	pedal hi-hat
		OHH	open hi-hat
		TB	tambourine
	RD	RD	ride cymbal
		RB	ride bell
	BE	CB	cowbell
		CRC	crash cymbal
	CY	SPC	splash cymbal
		CHC	Chinese cymbal
	CL	CL	clave/sticks



# MORE DRUM INSTRUMENTS?

## ■ Natural imbalance of data

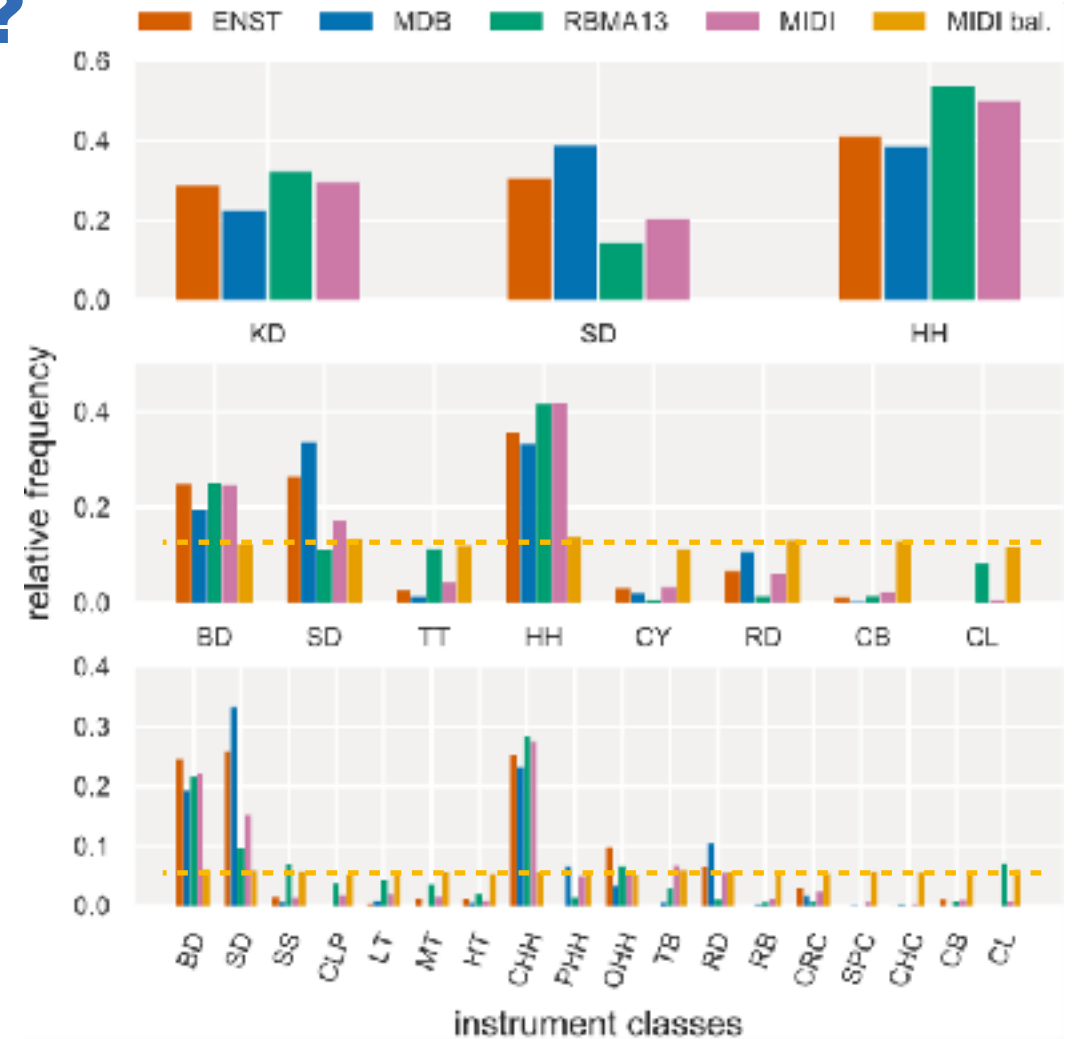
- ▶ Some instruments are used sparsely
- ▶ Few samples for those instruments
- ▶ Problem during NN training
- ▶ Problem for evaluation

## ■ Create synthetic dataset!

- ▶ ~4000 tracks
- ▶ More suitable sample

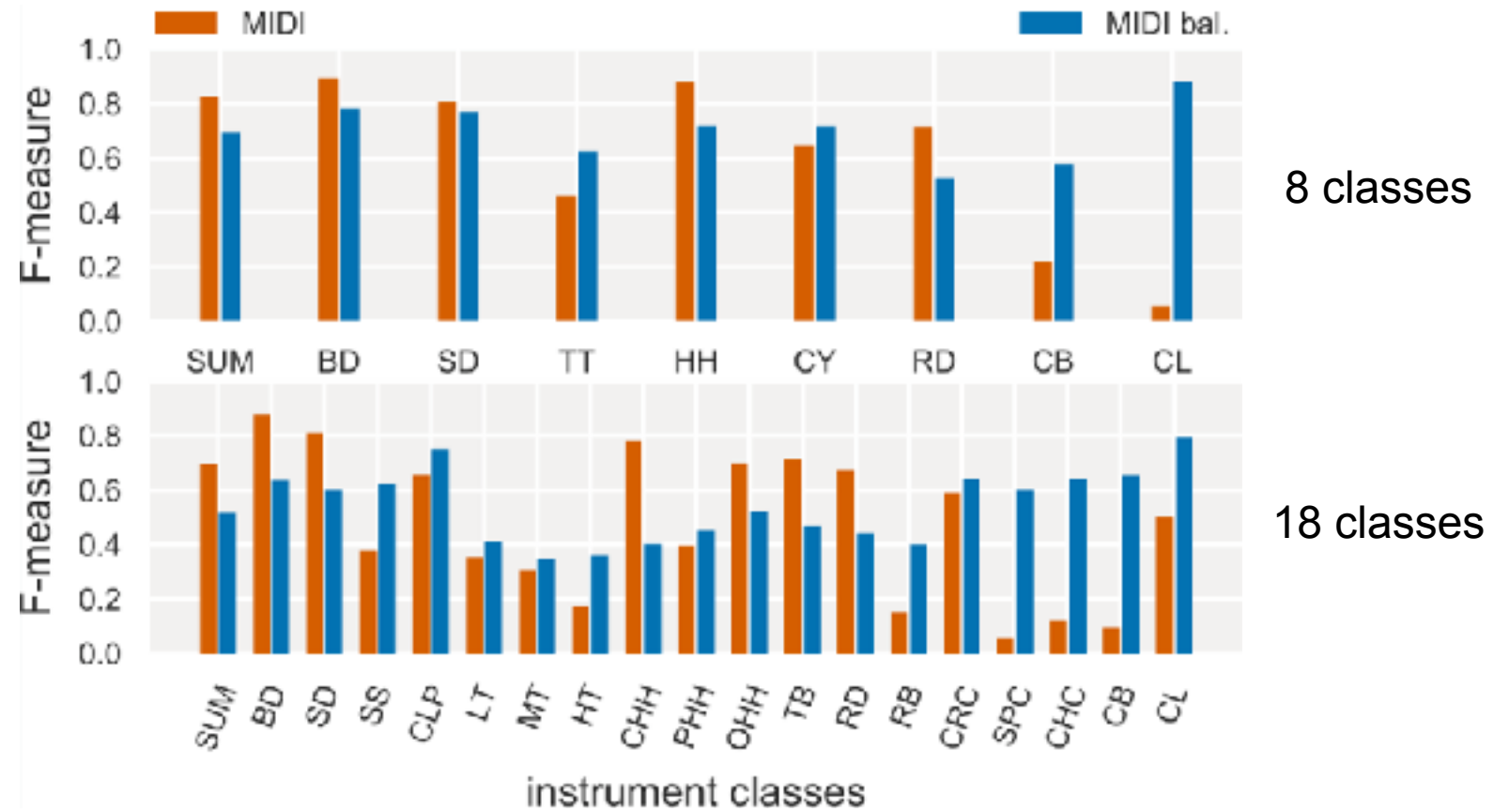
## ■ Balance instruments?

- ▶ All instruments equally represented 👍
- ▶ Artificial drum patterns 😞



Distribution of drum instruments in datasets

# PERFORMANCE ON SYNTHETIC DATA



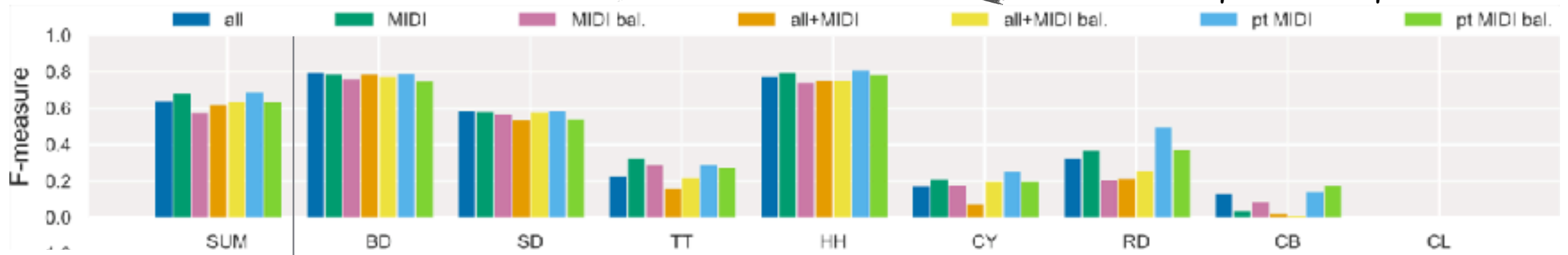
# PERFORMANCE ON REAL DATA

bar color = dataset used for training

*trained on mix of public datasets*

*bal. = balanced classes*

*pt = using pre-training*



*overall performance*

**CRNN with 8 classes on ENST**

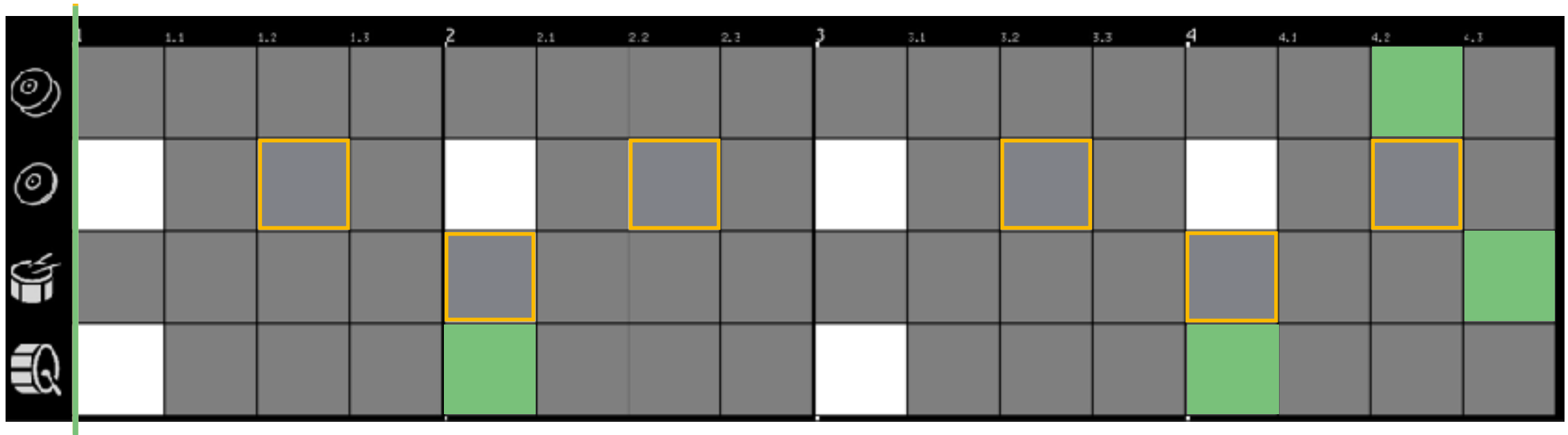
# CONCLUSIONS PART 1

- Improve drum transcription performance using **CRNN** models
- **Data augmentation** can be helpful
- **Multi-task learning** for drums and beats can be beneficial for recurrent architectures
- For more instruments: **pre-training** on large synthetic dataset

# PART 2

# AUTOMATIC DRUM PATTERN VARIATION

# WHAT IS DRUM PATTERN VARIATION?



- Create modifications of a given seed pattern
- Maintain characteristic of the beat
- **Add details** to increase intensity
- **Remove onsets** to make it more simple

# WHY AUTOMATIC DRUM PATTERN VARIATION?

- As an inspirational tool
- Increase productivity
- Exploration and experimentation
  
- Use cases
  - ▶ Music production (digital studio)
  - ▶ Live performances (experimental music)
  
- Challenges
  - ▶ Many degrees of freedom
  - ▶ Genre dependent
  - ▶ Original, meaningful, but not random patterns!



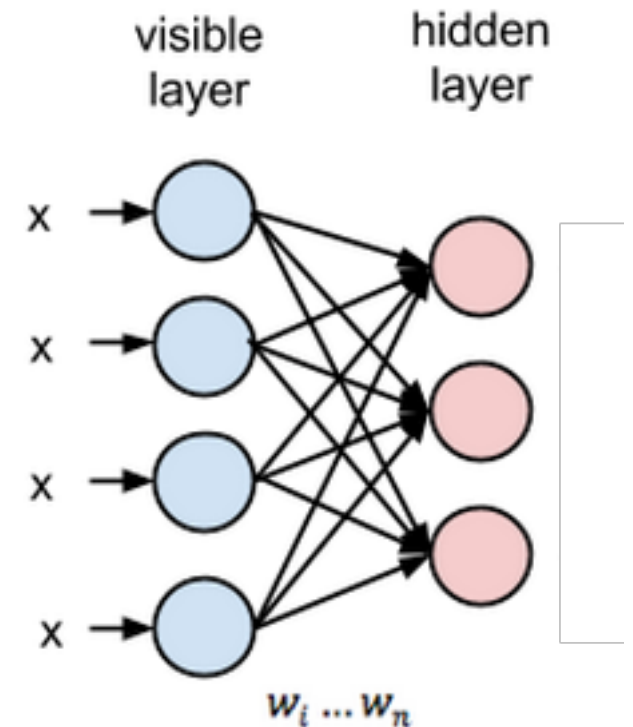
Reactable ROTOR



NI Maschine

# METHOD

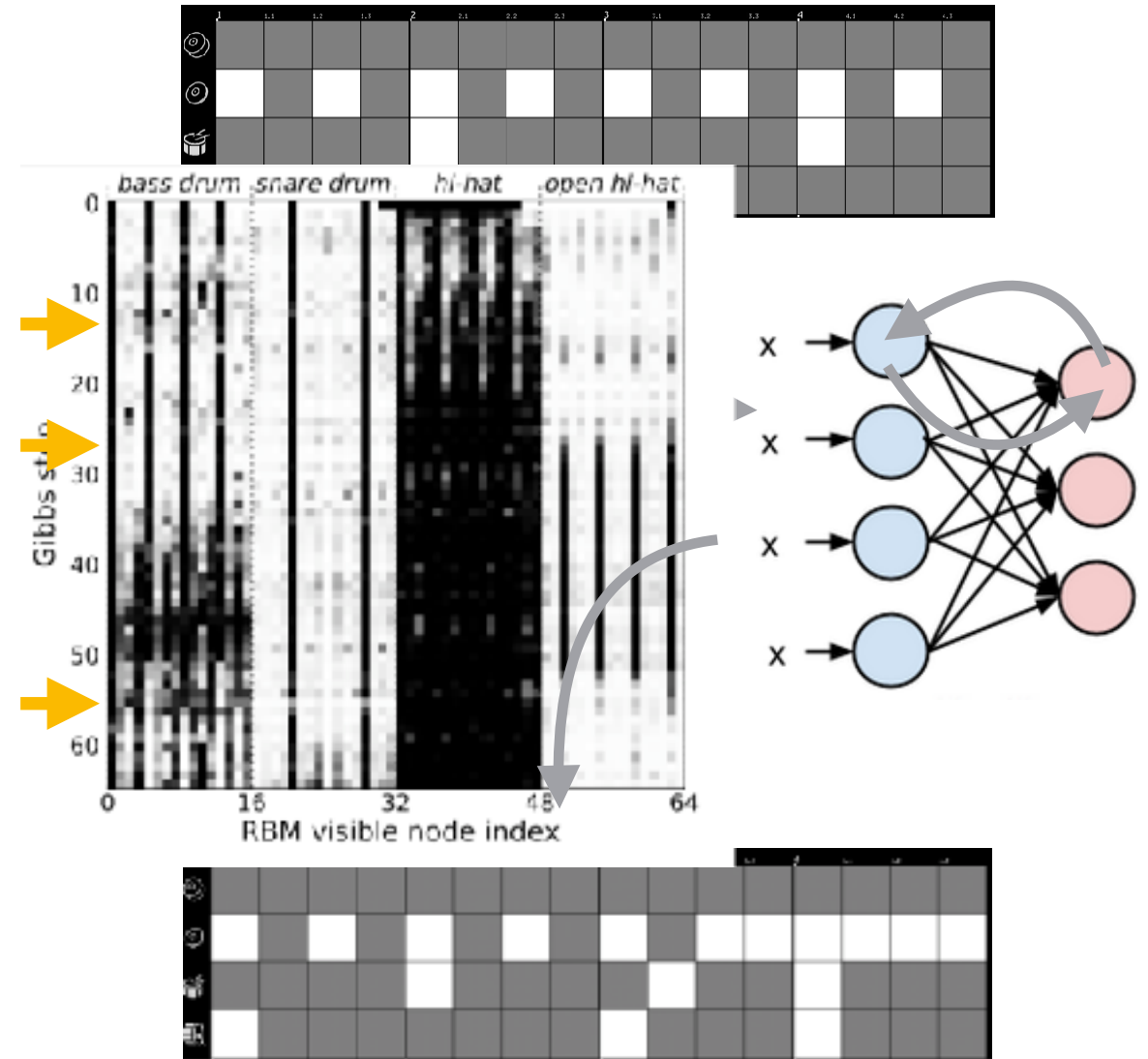
- Focus on EDM
- Step Sequencer Interface (4/4 time signature, 16<sup>th</sup> note resolution)
  - ▶ Fixed pattern grid size
- Stochastic generative model
- Seed pattern
  - ▶ Defines genre / style
  - ▶ Baseline for sorting of patterns
- ▶ Sampling of Restricted Boltzmann machine (RBM)
- ▶ Train on EDM drum loop library (NI Maschine)





# VARIATION METHOD

- Train RBM using drum loop database
- To create variations:
  - ▶ Enter seed pattern
  - ▶ Perform Gibbs sampling steps
  - ▶ Select and sort generated patterns
  - ▶ Provide patterns as variations



# DRUM PATTERN VARIATION - UI PROTOTYPES



# EVALUATION

## ■ Qualitative user studies for both UI prototypes

- ▶ Different pattern variation implementations

Richard Vogl and Peter Knees, “An Intelligent Drum Machine for Electronic Dance Music Production and Performance,” in *Proc. 17th Intl. Conf. for New Interfaces for Musical Expression (NIME)*, Copenhagen, DK, May 2017.

## ■ Quantitative survey for different pattern variation methods

- ▶ Database lookup based
- ▶ Genetic algorithm
- ▶ RBM based variation

R. Vogl, M. Leimeister, C. Ó Nuanáin, S. Jordà, M. Hlatky, and P. Knees, “An Intelligent Interface for Drum Pattern Variation and Comparative Evaluation of Algorithms,” *Journal of the Audio Engineering Society*, Vol. 64, No. 7, July 2016.

## An Intelligent Drum Machine for Electronic Dance Music Production and Performance

Richard Vogl<sup>1</sup> (richard.vogl@tuwien.ac.at)

Peter Knees<sup>1</sup> (peter.knees@tuwien.ac.at)



**introduction**

- ▶ drum tracks for electronic dance music (EDM) arrangements are usually built by repetition and variation of drum patterns
- ▶ patterns are manually created or unrefined patterns from a pattern library are used
- ▶ manual composition and creative variation is a labor intensive task but often the preferred method
- ▶ in this work an intelligent system is proposed to improve the workflow for the task at hand
- ▶ a goal is to make the process fun and spark creativity

**prototype**

- ▶ three steps: sequence generation by order-based pattern search, movable pattern variables
- ▶ touch screen based, touch-based interaction
- ▶ controls for playback, pattern selection, as well as undo and redo are part of the UI
- ▶ MIDI input evaluation: to ease integration with synthesizers

**variation algorithm**

- ▶ an iterative method as a global search is the focus of a reduced feature space (RFS) is used
- ▶ the RFS is based on a data set of 1,700 one bar EDM drum patterns
- ▶ to create variations the used pattern is stored in the hidden layer of the RBM and varied using Gibbs sampling

**evaluation**

- ▶ qualitative elements with five experts
- ▶ specific aspects evaluated using a questionnaire using five-point Likert scales (see evaluation slide)
- ▶ participants were exploring the prototype and comparing it to the previous, feature-rich database-based version
- ▶ two different variation algorithms were evaluated

**conclusion**

- ▶ participants found the interface to be intuitive, usable
- ▶ an aspect of the interface was less used compared to the previous prototype
- ▶ consistency of patterns improved, but only at the expense of the usability to generate files
- ▶ the additional features were received positively
- ▶ all of the touch interface was generally received as an improvement, except areas for physical knobs on the touch interface was low

**evaluation results table:**

algorithm	#		non-RBM		RBM
	U	S	mean	std	
usability	4.80	3.20	4.70	0.50	5
usability	4.40	4.20	4.60	0.10	2
efficiency	3.80	3.20	3.70	0.10	2
diversity	3.80	3.20	3.80	0.10	2
interestingness	4.80	3.20	4.40	0.40	2
usability	4.40	3.20	4.40	0.10	2
div	3.80	4.20	4.00	0.20	2

**non-RBM table:**

	U	S	mean	std
usability	4.70	0.50	4.70	0.50
div	4.80	3.20	4.00	0.50
efficiency	3.70	0.10	3.70	0.10

**total table:**

	U	S	mean	std
usability	4.70	0.50	4.70	0.50
div	4.80	3.20	4.00	0.50
efficiency	3.70	0.10	3.70	0.10

# DEMO

# IN PROGRESS: DRUM PATTERN GENERATION

- Input parameters
  - ▶ Music style
  - ▶ Intensity/loudness
  - ▶ Complexity
- More Instruments
- Higher time resolution
- Collect training data using drum transcription
- Generative adversarial networks (GANs)



*Apple Logic Pro X: Drummer*

# VISION: AUTOMATIC DRUMMER?

- Combine everything to build an **fully automatic drummer**?
  - ▶ Better **drum transcription** for large volume of **training** examples
  - ▶ Integrate more powerful models for **pattern generation**
  - ▶ Apply other MIR techniques to **identify genre** and **follow the beat**