DRUM TRANSCRIPTION VIA JOINT BEAT AND DRUM MODELING USING CONVOLUTIONAL RNNs

Richard Vogl\textsuperscript{1,2}, Matthias Dorfer\textsuperscript{2}, Gerhard Widmer\textsuperscript{2}, Peter Knees\textsuperscript{1}
richard.vogl@tuwien.ac.at, matthias.dorfer@jku.at, gerhard.widmer@jku.at, peter.knees@tuwien.ac.at
WHAT IS DRUM TRANSCRIPTION?

- **Input:** western popular music containing drums
- **Output:** symbolic representation of notes played by drum instruments
WHAT IS DRUM TRANSCRIPTION?

Focus on the three major drum instruments:

- bass or kick drum (KD)
- snare drum (SD)
- hi-hat (HH)

Reasons:

- Dominant instruments: most onsets
- Common subset for public datasets
WHY DRUM TRANSCRIPTION?
WHY DRUM TRANSCRIPTION?

- Wide range of application for transcripts:
WHY DRUM TRANSCRIPTION?

- Wide range of application for transcripts:
  - Generate **sheet music**
WHY DRUM TRANSCRIPTION?

- Wide range of application for transcripts:
  - Generate **sheet music**
  - **Music production**
    - sampling / remixing / resynthesis
**WHY DRUM TRANSCRIPTION?**

- Wide range of application for transcripts:
  - Generate **sheet music**
  - **Music production**
    - sampling / remixing / resynthesis
  - Higher level **MIR tasks**
    - use drum patterns for other tasks
    - genre classification
    - song segmentation
STATE OF THE ART
STATE OF THE ART

Overview Article

STATE OF THE ART

- Overview Article

- Current state-of-the-art systems:
  - End-to-end / activation-function-based approaches
  - NMF based approaches and RNN approaches

![spectrogram](image.png)  ![activation functions](image.png)
SYSTEM OVERVIEW

audio → signal preprocessing → NN feature extraction event detection classification → peak picking → events

NN training → database
SYSTEM OVERVIEW

1. Audio
2. Signal preprocessing
3. Feature extraction
4. Event detection
5. Classification
6. Peak picking
7. NN training
8. Events

Audio events spectrogram
SYSTEM OVERVIEW

Audio signals undergo preprocessing, feature extraction, event detection, and classification. The neural network (NN) is trained on these features, and peak picking is used to identify events. Activation functions are applied to the classified data. The system involves signal processing, audio events, spectrograms, and frequency-time representations.
SYSTEM OVERVIEW

Audio signals are preprocessed for feature extraction, which includes event detection and classification. Afterward, peak picking is applied. The extracted features are used for NN training, leading to activation functions for generating events.
ISSUES OF CURRENT SYSTEMS
ISSUES OF CURRENT SYSTEMS

- Performance not satisfying on real music
ISSUES OF CURRENT SYSTEMS

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

- Performance not satisfying on real music
- Do not produce additional information for transcripts
  
  \textit{drum onset detection} vs \textit{drum transcription}

  \begin{itemize}
  \item bars lines
  \end{itemize}
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
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
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
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
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
- Only three instrument classes
- etc.
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
- Only three instrument classes
- etc.
ADDITIONAL INFORMATION FOR TRANSCRIPTS
ADDITIONAL INFORMATION FOR TRANSCRIPTS
Use **beat and downbeat tracking** to get:

- bars lines
- tempo
- meter
Use **beat and downbeat tracking** to get:

- bars lines
- tempo
- meter
Use **beat and downbeat tracking** to get:

- bars lines
- tempo
- meter
IMPROVE PERFORMANCE

Three components to reach this goal:

1. Leverage beat information
2. Better model for drum detection
3. Dataset with real music for training
1. LEVERAGE BEAT INFORMATION
1. LEVERAGE BEAT INFORMATION

Beats are **highly correlated** with drum patterns
1. 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).
1. 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
MULTI-TASK LEARNING

input

output

$f [\text{Hz}]$

$t [s]$
MULTI-TASK LEARNING

Three experiments:
MULTI-TASK LEARNING

Three experiments:
- Training on drum targets ($DT$)
MULTI-TASK LEARNING

Three experiments:

- Training on drum targets ($DT$)
- Training on drum targets with annotated beats as additional input features ($BF$)
MULTI-TASK LEARNING

Three experiments:

- Training on drum targets ($DT$)
- Training on drum targets with annotated beats as additional input features ($BF$)
- Training on drum and beat targets as multi-task problem ($MT$)
MULTI-TASK LEARNING

Three experiments:
- Training on drum targets \((DT)\)
- Training on drum targets with annotated beats as additional input features \((BF)\)
- Training on drum and beat targets as multi-task problem \((MT)\)

Expected increase in performance for \(BF\) compared to \(DT\)
MULTI-TASK LEARNING

Three experiments:
- Training on drum targets ($DT$)
- Training on drum targets with annotated beats as additional input features ($BF$)
- Training on drum and beat targets as multi-task problem ($MT$)

Expected increase in performance for $BF$ compared to $DT$
Expected increase in performance for $MT$ compared to $DT$
2. NETWORK MODELS — BASELINE MODELS
2. NETWORK MODELS — BASELINE MODELS

- Recurrent neural networks
2. NETWORK MODELS — BASELINE MODELS

- Recurrent neural networks
  - Recurrent connections act as **memory**
  - Processing of **sequential data**

RNN train data sample
### 2. NETWORK MODELS — BASELINE MODELS

- **Recurrent neural networks**
  - Recurrent connections act as memory
  - Processing of sequential data
  - Work well for drum detection and beat tracking

[Boek et al. ISMIR’16]

![RNN train data sample](image)
2. NETWORK MODELS — BASELINE MODELS

- Recurrent neural networks
  - Recurrent connections act as memory
  - Processing of sequential data
  - Work well for drum detection and beat tracking
    [Böck et al. ISMIR’16]

- RNN with label time shift (tsRNN)
  state-of-the-art baseline [Vogl et al. ICASSP’17]

- Bidirectional recurrent NN (BDRNN)
  [Vogl et al. ISMIR’16] [Southall et al. ISMIR’16]
  - Similar performance tsRNN

RNN train data sample
2. NETWORK MODELS — NEW FOR DT
2. NETWORK MODELS — NEW FOR DT

- Convolutional NN (CNN)
  - Convolutions capture local correlations
  - Acoustic modeling of drum sounds
2. NETWORK MODELS — NEW FOR DT

- 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
## NETWORK MODELS

<table>
<thead>
<tr>
<th></th>
<th>Frames</th>
<th>Context</th>
<th>Conv. Layers</th>
<th>Rec. Layers</th>
<th>Dense Layers</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNN (S)</td>
<td>100</td>
<td>—</td>
<td>—</td>
<td>2x50 GRU</td>
<td>—</td>
</tr>
<tr>
<td>RNN (L)</td>
<td>400</td>
<td>—</td>
<td>—</td>
<td>3x30 GRU</td>
<td>—</td>
</tr>
<tr>
<td>CNN (S)</td>
<td>—</td>
<td>9</td>
<td>2 x 32 3x3 filt. 3x3 max pooling</td>
<td>—</td>
<td>2x256</td>
</tr>
<tr>
<td>CNN (L)</td>
<td>—</td>
<td>25</td>
<td>2 x 64 3x3 filt. 3x3 max pooling</td>
<td>—</td>
<td>2x256</td>
</tr>
<tr>
<td>CRNN (S)</td>
<td>100</td>
<td>9</td>
<td>3x3 max pooling all w/ batch norm.</td>
<td>2x50 GRU</td>
<td>—</td>
</tr>
<tr>
<td>CRNN (L)</td>
<td>400</td>
<td>13</td>
<td>3x60 GRU</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>tsRNN</td>
<td>state-of-the-art baseline [Vogl et al. ICASSP’17]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
CLASSIC DATASETS (ONLY DRUMS)
CLASSIC DATASETS (ONLY DRUMS)

- **IDMT-SMT-Drums** [Dittmar and Gärtner 2014]
  - Solo drum tracks, recorded, synthesized, and sampled
  - 95 tracks, total: **24m**, onsets: **8004** + training samples
CLASSIC DATASETS (ONLY DRUMS)

- **IDMT-SMT-Drums** [Dittmar and Gärtner 2014]
  - Solo drum tracks, recorded, synthesized, and sampled
  - 95 tracks, total: 24m, onsets: 8004 + training samples
CLASSIC DATASETS (ONLY DRUMS)

- **IDMT-SMT-Drums** [Dittmar and Gärtner 2014]
  - Solo drum tracks, recorded, synthesized, and sampled
  - 95 tracks, total: **24m**, onsets: **8004** + training samples

- **ENST-Drums** [Gillet and Richard 2006]
  - Recordings, three drummers on different drum kits, **optional accompaniment**
  - 64 tracks, total: **1h**, onsets: **22391** + training samples
CLASSIC DATASETS (ONLY DRUMS)

- **IDMT-SMT-Drums** [Dittmar and Gärtner 2014]
  - Solo drum tracks, recorded, synthesized, and sampled
  - 95 tracks, total: **24m**, onsets: **8004** + training samples

- **ENST-Drums** [Gillet and Richard 2006]
  - Recordings, three drummers on different drum kits, **optional accompaniment**
  - 64 tracks, total: **1h**, onsets: **22391** + training samples
CLASSIC DATASETS (ONLY DRUMS)

- **IDMT-SMT-Drums** [Dittmar and Gärtner 2014]
  - Solo drum tracks, recorded, synthesized, and sampled
  - 95 tracks, total: 24m, onsets: 8004 + training samples

- **ENST-Drums** [Gillet and Richard 2006]
  - Recordings, three drummers on different drum kits, **optional accompaniment**
  - 64 tracks, total: 1h, onsets: 22391 + training samples
EXPERIMENTS PART 1
EXPERIMENTS PART 1

- Evaluation of our six models on classic datasets for drum detection (DT)
EXPERIMENTS PART 1

- Evaluation of our six models on **classic datasets** for **drum detection** (DT)

- **SMT solo**
  - Three-fold cross-validation on the three different types of solo drum tracks
EXPERIMENTS PART 1

Evaluation of our six models on classic datasets for drum detection (DT)

- SMT solo
  - Three-fold cross-validation on the three different types of solo drum tracks

- ENST solo
  - Three-fold cross-validation on solo drum tracks of the three different drummers
EXPERIMENTS PART 1

Evaluation of our six models on classic datasets for drum detection (DT)

SMT solo
- Three-fold cross-validation on the three different types of solo drum tracks

ENST solo
- Three-fold cross-validation on solo drum tracks of the three different drummers

ENST acc.
- Three-fold cross-validation on tracks with accompaniment
DT 3-FOLD CV RESULTS ON CLASSIC DATASETS

- **F-measure [%]**
- **Baseline**
- **RNN (S)**
- **RNN (L)**
- **CNN (S)**
- **CNN (L)**
- **CRNN (S)**
- **CRNN (L)**

### Results by Dataset:
- **SMT solo**
- **ENST solo**
- **ENST acc.**
3. NEW DATASETS (DRUMS AND BEATS)

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
3. NEW DATASETS (DRUMS AND BEATS)

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
3. NEW DATASETS (DRUMS AND BEATS)

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
3. NEW DATASETS (DRUMS AND BEATS)

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

**Multi-task** evaluation

- **DT**: Drum transcription / three fold cross-validation (same as on SMT and ENST)
- **BF**: Drum transcription using annotated beats as additional input features
- **MT**: Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13

- DT
- BF
- MT

F-measure [%]
- 50
- 55
- 60
- 65
- 70

RNN (S)
RNN (L)
CNN (S)
CNN (L)
CRNN (S)
CRNN (L)
RESULTS ON RBMA13

The diagram shows the F-measure percentages for different methods on DT, BF, and MT tasks. The methods include RNN (S), RNN (L), CNN (S), CNN (L), CRNN (S), and CRNN (L). The F-measures range from 50% to 70% with a clear differentiation between the methods and tasks.
RESULTS ON RBMA13: RNNs

DT … Drum transcription (3-fold CV)
BF … Drum transcription using annotated beats as additional input features
MT … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: RNNs

Impact on RNNs:

- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: RNNs

Impact on RNNs:
- BF improves for both models ✓
RESULTS ON RBMA13: RNNs

Impact on RNNs:
- BF improves for both models ✔
- MT improves for both models ✔
RESULTS ON RBMA13: RNNs

Impact on RNNs:
- BF improves for both models ✓
- MT improves for both models ✓
- MT even better than BF for small model !

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>F-measure [%]</th>
<th>RNN (S)</th>
<th>RNN (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>64</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>MT</td>
<td>66</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>

Legend:
- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: CNNs

- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: CNNs

Impact on CNNs:

- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning

![Bar chart showing F-measure [%] for different methods.](image)

- CNN (S) with various impacts on its performance.
- CNN (L) with corresponding impacts.

F-measure [%] values range from 50 to 70.
RESULTS ON RBMA13: CNNs

Impact on CNNs:
- BF inconsistent
RESULTS ON RBMA13: CNNs

Impact on CNNs:
- BF inconsistent
- MT declines for both models

<table>
<thead>
<tr>
<th>Model</th>
<th>BF Consistent</th>
<th>MT Decline</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN (S)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CNN (L)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: CRNNs

- **CRNN (S)**
- **CRNN (L)**

<table>
<thead>
<tr>
<th>Method</th>
<th>F-measure [%]</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT</td>
<td>65</td>
</tr>
<tr>
<td>BF</td>
<td>66</td>
</tr>
<tr>
<td>MT</td>
<td>67</td>
</tr>
</tbody>
</table>

- **DT** … Drum transcription (3-fold CV)
- **BF** … Drum transcription using annotated beats as additional input features
- **MT** … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: CRNNs

Impact on CRNNs:

- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: CRNNs

Impact on CRNNs:

- BF improves for both models ✔

Impact on CRNNs:
- BF improves for both models ✔

- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: CRNNs

Impact on CRNNs:
- BF improves for both models ✔
- MT improves for small models ✔

Impact on CRNNs:

<table>
<thead>
<tr>
<th>Method</th>
<th>CRNN (S)</th>
<th>CRNN (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DT … Drum transcription (3-fold CV)</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>BF … Drum transcription using annotated beats as additional input features</td>
<td>✔</td>
<td>✔</td>
</tr>
<tr>
<td>MT … Drum transcription and beat detection via multi-task learning</td>
<td>✔</td>
<td>✔</td>
</tr>
</tbody>
</table>
RESULTS ON RBMA13: CRNNs

Impact on CRNNs:
- BF improves for both models ✓
- MT improves for small models ✓
- MT even better than BF for small model !

<table>
<thead>
<tr>
<th>Method</th>
<th>CRNN (S)</th>
<th>CRNN (L)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BF</td>
<td>65</td>
<td>65</td>
</tr>
<tr>
<td>MT</td>
<td>68</td>
<td>68</td>
</tr>
</tbody>
</table>

DT … Drum transcription (3-fold CV)
BF … Drum transcription using annotated beats as additional input features
MT … Drum transcription and beat detection via multi-task learning
RESULTS ON RBMA13: CRNNs

Impact on CRNNs:
- BF improves for both models ✔
- MT improves for small models ✔
- MT even better than BF for small model !
- MT equal for large model ?

- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning
RESULTS FOR RECURRENT ARCHITECTURES

- **F-measure [%]**
  - 50
  - 55
  - 60
  - 65
  - 70

- **Architectures**
  - RNN (S)
  - RNN (L)
  - CRNN (S)
  - CRNN (L)

- **Techniques**
  - DT … Drum transcription (3-fold CV)
  - BF … Drum transcription using annotated beats as additional input features
  - MT … Drum transcription and beat detection via multi-task learning
RESULTS FOR RECURRENT ARCHITECTURES

- **RNN (S)**
- **RNN (L)**
- **CRNN (S)**
- **CRNN (L)**

- **DT**: Drum transcription (3-fold CV)
- **BF**: Drum transcription using annotated beats as additional input features
- **MT**: Drum transcription and beat detection via multi-task learning
RESULTS FOR RECURRENT ARCHITECTURES
RESULTS FOR RECURRENT ARCHITECTURES

No improvement because of beat tracking results?
RESULTS RECAP
RESULTS RECAP

- Both CRNN (S/L) and CNN (L) perform **better than the baselines** for classic datasets ✔
RESULTS RECAP

- Both CRNN (S/L) and CNN (L) perform better than the baselines for classic datasets ✔
- BF is always better than DT for (C)RNN ✔
RESULTS RECAP

- Both CRNN (S/L) and CNN (L) perform **better than the baselines** for classic datasets ✔
- **BF is always better** than DT for (C)RNN ✔
- **MT is always better** than DT for (C)RNN except for CRNN (L) ✔ / ?
  - Possibly due to bad beat tracking results
RESULTS RECAP

- Both CRNN (S/L) and CNN (L) perform **better than the baselines** for classic datasets ✔
- **BF is always better** than DT for (C)RNN ✔
- **MT is always better** than DT for (C)RNN except for CRNN (L) ✔ / ?
  - Possibly due to bad beat tracking results
- **MT is better than BF** for small (C)RNN ✔
RESULTS RECAP

- Both CRNN (S/L) and CNN (L) perform **better than the baselines** for classic datasets ✔
- **BF is always better** than DT for (C)RNN ✔
- **MT is always better** than DT for (C)RNN except for CRNN (L) ✔ / ？
  - Possibly due to bad beat tracking results
- **MT is better than BF** for small (C)RNN ✔
- Beats are **not beneficial for CNNs** !
### MIREX’17 RESULTS

<table>
<thead>
<tr>
<th>Algorithm</th>
<th>mean fm</th>
<th>mean pr</th>
<th>mean rc</th>
<th>BD mean fm</th>
<th>SD mean fm</th>
<th>HH mean fm</th>
</tr>
</thead>
<tbody>
<tr>
<td>CW1</td>
<td>0.51</td>
<td>0.46</td>
<td>0.68</td>
<td>0.68</td>
<td>0.48</td>
<td>0.38</td>
</tr>
<tr>
<td>CW3</td>
<td>0.53</td>
<td>0.50</td>
<td>0.65</td>
<td>0.67</td>
<td>0.46</td>
<td>0.42</td>
</tr>
<tr>
<td>CW2</td>
<td>0.55</td>
<td>0.52</td>
<td>0.66</td>
<td>0.70</td>
<td>0.55</td>
<td>0.40</td>
</tr>
<tr>
<td>RV3</td>
<td>0.68</td>
<td>0.74</td>
<td>0.70</td>
<td>0.81</td>
<td>0.64</td>
<td>0.51</td>
</tr>
<tr>
<td>RV2</td>
<td>0.67</td>
<td>0.69</td>
<td>0.73</td>
<td>0.78</td>
<td>0.67</td>
<td>0.51</td>
</tr>
<tr>
<td>RV1</td>
<td>0.71</td>
<td>0.75</td>
<td>0.74</td>
<td>0.82</td>
<td>0.70</td>
<td>0.53</td>
</tr>
<tr>
<td>RV4</td>
<td>0.70</td>
<td>0.74</td>
<td>0.73</td>
<td>0.81</td>
<td>0.70</td>
<td>0.52</td>
</tr>
<tr>
<td>CS1</td>
<td>0.61</td>
<td>0.56</td>
<td>0.73</td>
<td>0.79</td>
<td>0.55</td>
<td>0.46</td>
</tr>
<tr>
<td>CS3</td>
<td>0.63</td>
<td>0.59</td>
<td>0.75</td>
<td>0.78</td>
<td>0.58</td>
<td>0.49</td>
</tr>
<tr>
<td>CS2</td>
<td>0.63</td>
<td>0.61</td>
<td>0.71</td>
<td>0.78</td>
<td>0.57</td>
<td>0.49</td>
</tr>
</tbody>
</table>

MIREX’17 RESULTS

http://www.music-ir.org/mirex/wiki/2017:Drum_Transcription_Results
CONCLUSIONS

- Use beats and downbeats to get **meta information** for transcripts
CONCLUSIONS

- Use beats and downbeats to get meta information for transcripts
- Multi-task learning for drums and beats can be beneficial for recurrent architectures
CONCLUSIONS

- Use beats and downbeats to get meta information for transcripts
- Multi-task learning for drums and beats can be beneficial for recurrent architectures
- CRNNs can outperform RNNs
CONCLUSIONS

- Use beats and downbeats to get **meta information** for transcripts

- **Multi-task learning** for drums and beats can be beneficial for recurrent architectures

- **CRNNs** can outperform RNNs

- **CRNN best overall results** @ MIREX’17 drum transcription
  
  MIREX system: [http://ifs.tuwien.ac.at/~vogl/models/mirex-17.zip](http://ifs.tuwien.ac.at/~vogl/models/mirex-17.zip)
  
  madmom: [https://github.com/CPJKU/madmom](https://github.com/CPJKU/madmom)
CONCLUSIONS

- Use beats and downbeats to get **meta information** for transcripts
- **Multi-task learning** for drums and beats can be beneficial for recurrent architectures
- **CRNNs** can outperform RNNs

**CRNN best overall results** @ MIREX’17 drum transcription
MIREX system: [http://ifs.tuwien.ac.at/~vogl/models/mirex-17.zip](http://ifs.tuwien.ac.at/~vogl/models/mirex-17.zip)
madmom: [https://github.com/CPJKU/madmom](https://github.com/CPJKU/madmom)

- **New dataset** with free music featuring **beat**, and **drum annotations**
  [http://ifs.tuwien.ac.at/~vogl/datasets/](http://ifs.tuwien.ac.at/~vogl/datasets/)
CONCLUSIONS

- Use beats and downbeats to get **meta information** for transcripts

- **Multi-task learning** for drums and beats can be beneficial for recurrent architectures

- **CRNNs** can outperform RNNs

- **CRNN best overall results** @ MIREX’17 drum transcription
  MIREX system: [http://ifs.tuwien.ac.at/~vogl/models/mirex-17.zip](http://ifs.tuwien.ac.at/~vogl/models/mirex-17.zip)
  madmom: [https://github.com/CPJKU/madmom](https://github.com/CPJKU/madmom)

- **New dataset** with free music featuring **beat**, and **drum annotations**
  [http://ifs.tuwien.ac.at/~vogl/datasets/](http://ifs.tuwien.ac.at/~vogl/datasets/)