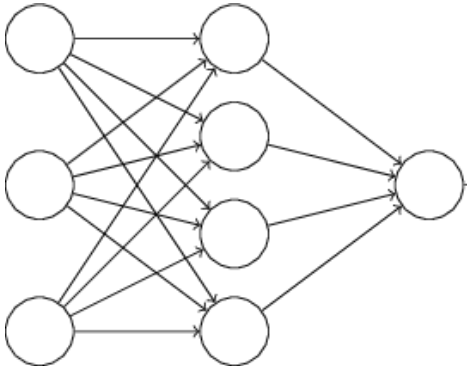
DRUM TRANSCRIPTION VIA JOINT BEAT AND DRUM MODELING USING CONVOLUTIONAL RNNs

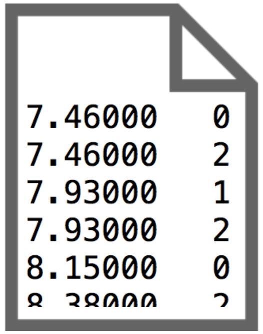
Richard Vogl[1,2], Matthias Dorfer[2], Gerhard Widmer[2], Peter Knees[1]
richard.vogl@tuwien.ac.at, matthias.dorfer@jku.at, gerhard.widmer@jku.at, peter.knees@tuwien.ac.at

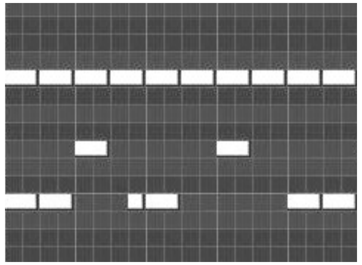
1 TU WIEN TECHNISCHE UNIVERSITÄT WIEN ifs mir group

2 JKU JOHANNES KEPLER UNIVERSITÄT LINZ Department of Computational Perception

WHAT IS DRUM TRANSCRIPTION?







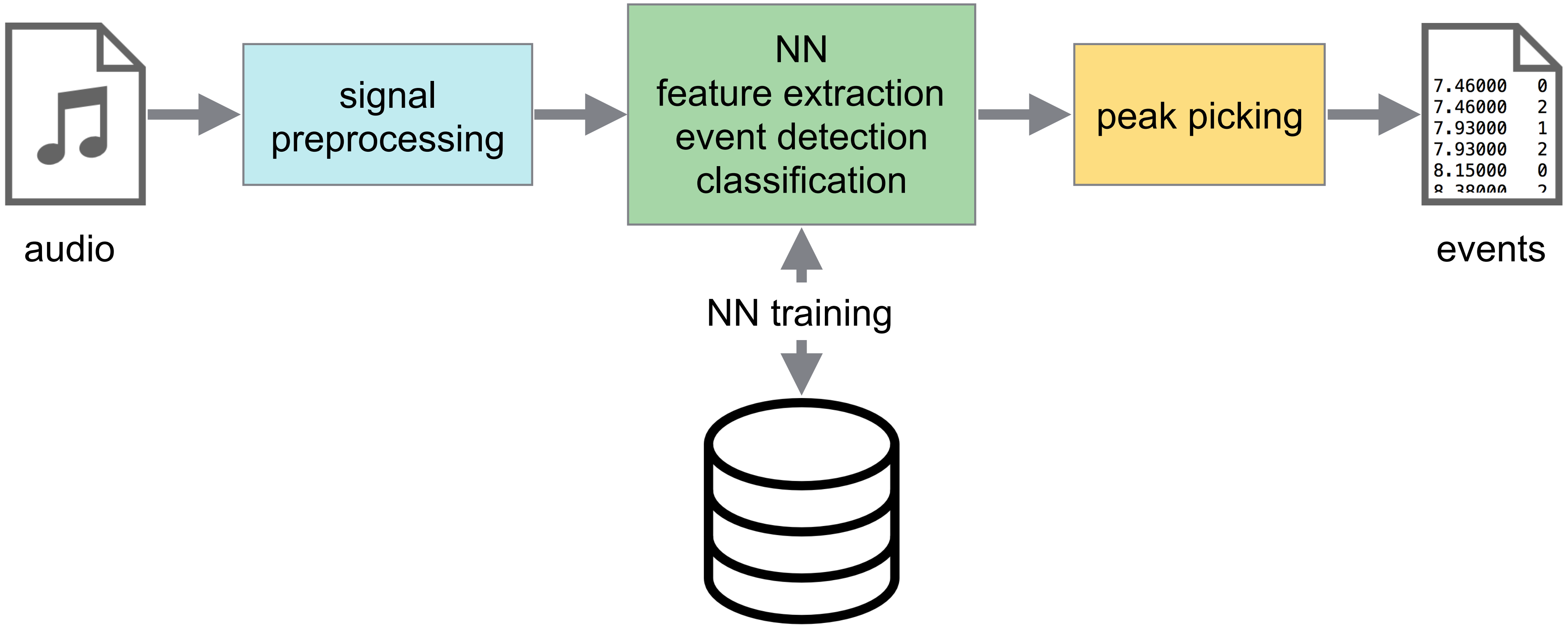
- **Input:** western popular music containing drums
- **Output:** symbolic representation of notes played by drum instruments

WHAT IS DRUM TRANSCRIPTION?

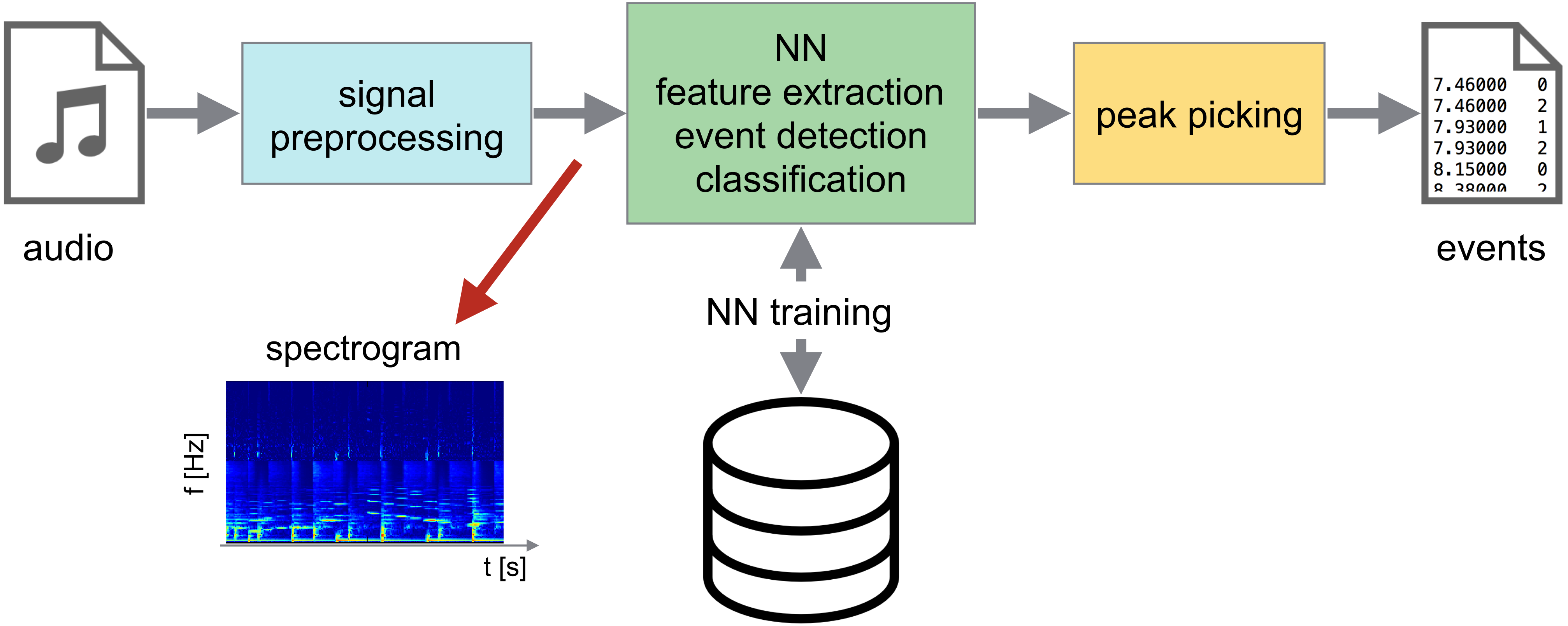
- Focus on the three major drum instruments:
  - bass or kick drum (**KD**)
  - snare drum (**SD**)
  - hi-hat (**HH**)

- Reasons:
  - Dominant instruments: most onsets
  - Common subset for public datasets

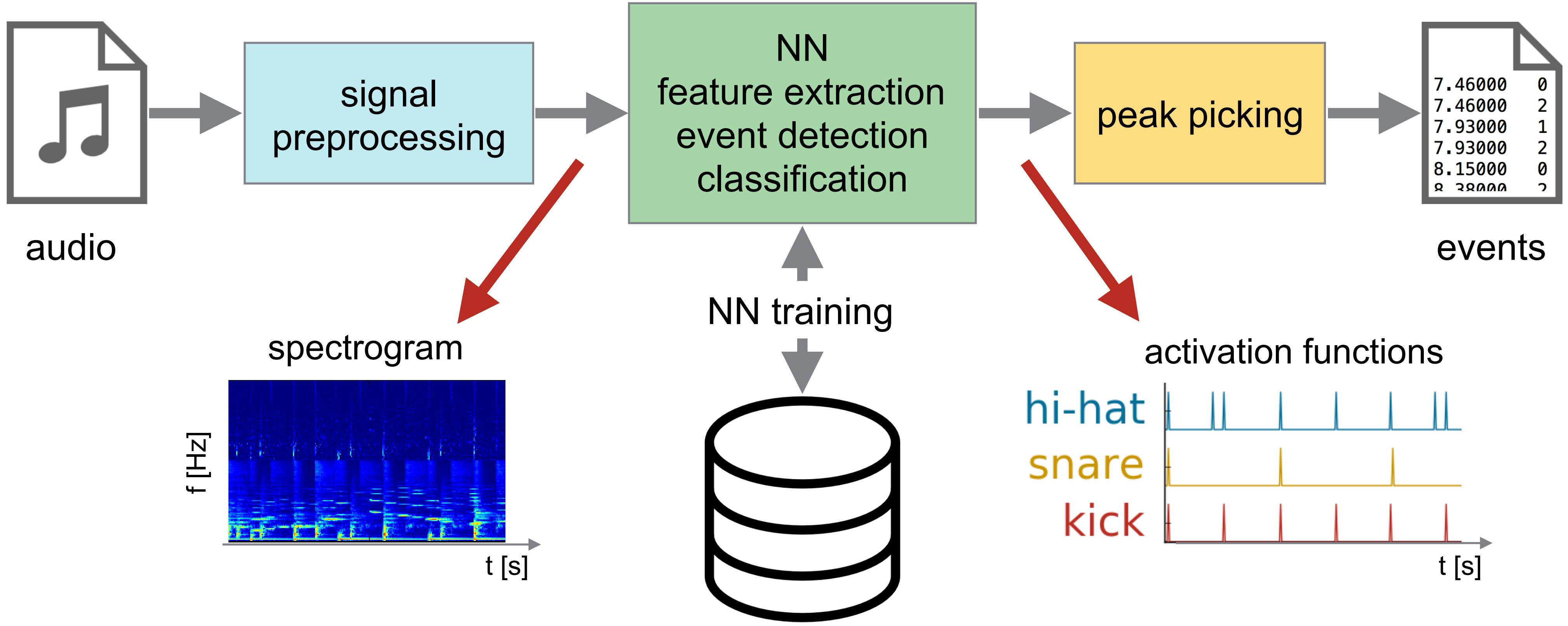
SYSTEM OVERVIEW



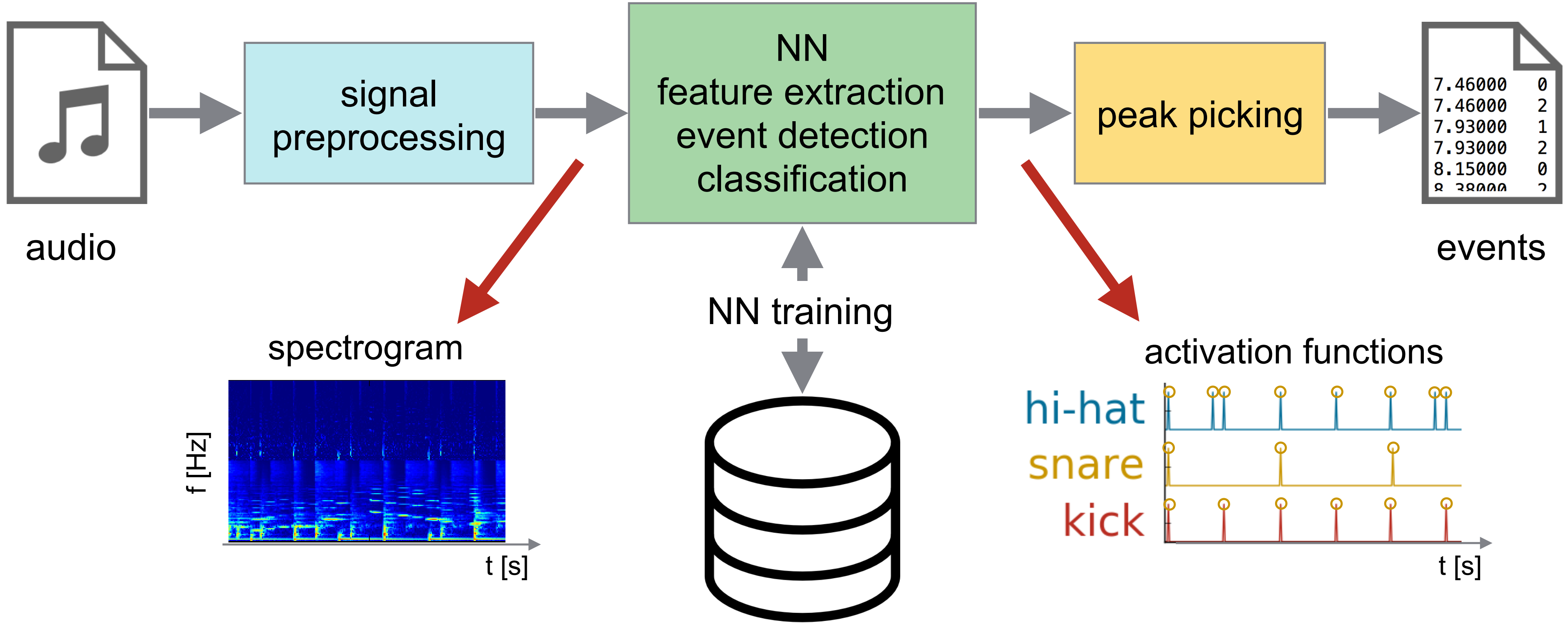
SYSTEM OVERVIEW



SYSTEM OVERVIEW



SYSTEM OVERVIEW



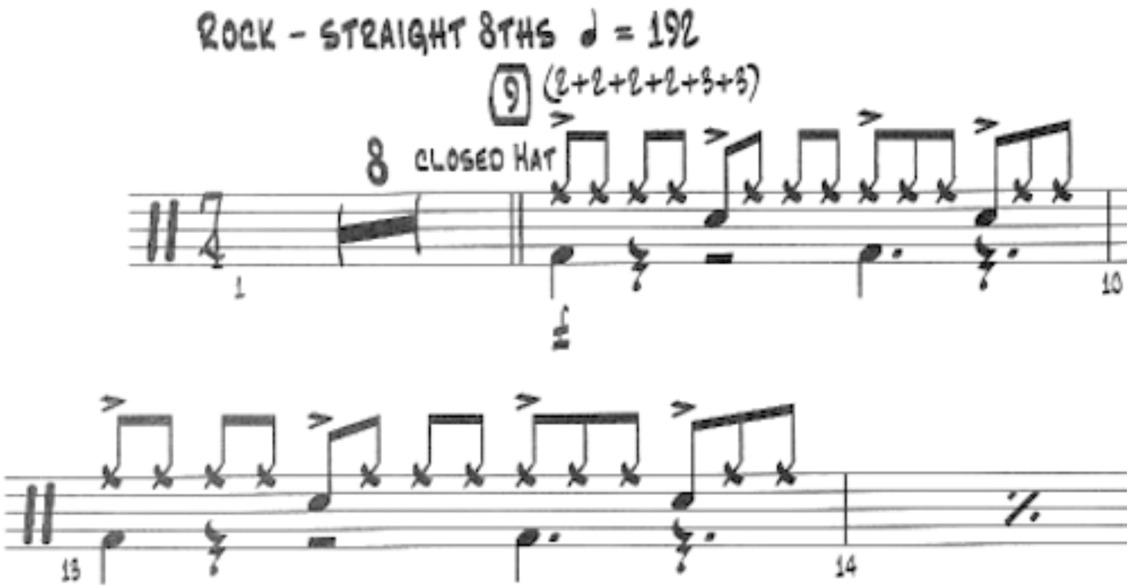
ISSUES OF CURRENT SYSTEMS

ISSUES OF CURRENT SYSTEMS

- Performance not satisfying on real music

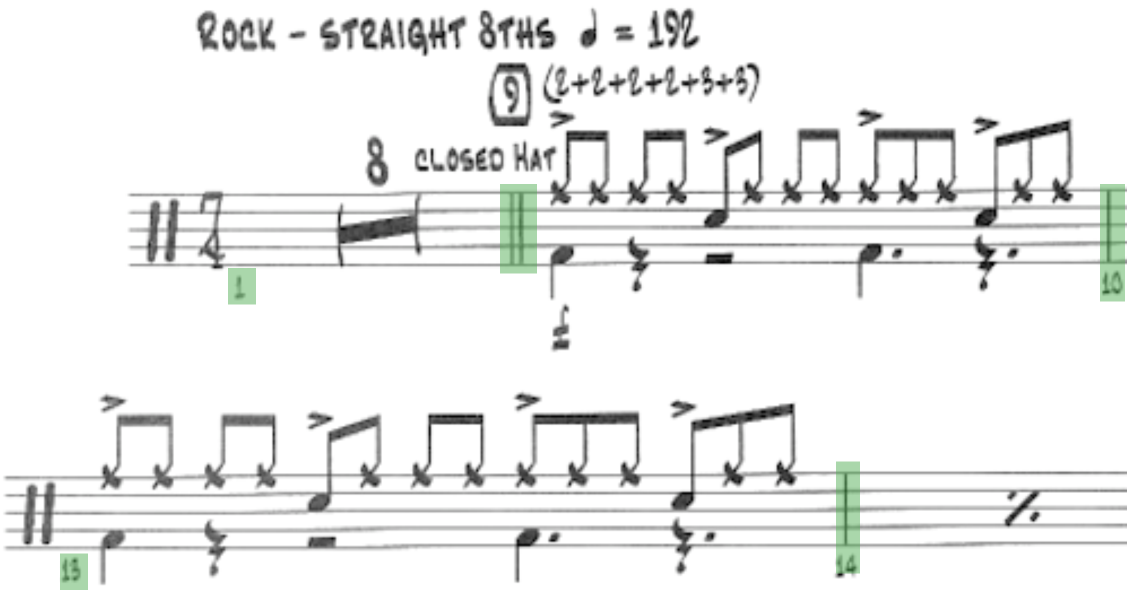
ISSUES OF CURRENT SYSTEMS

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



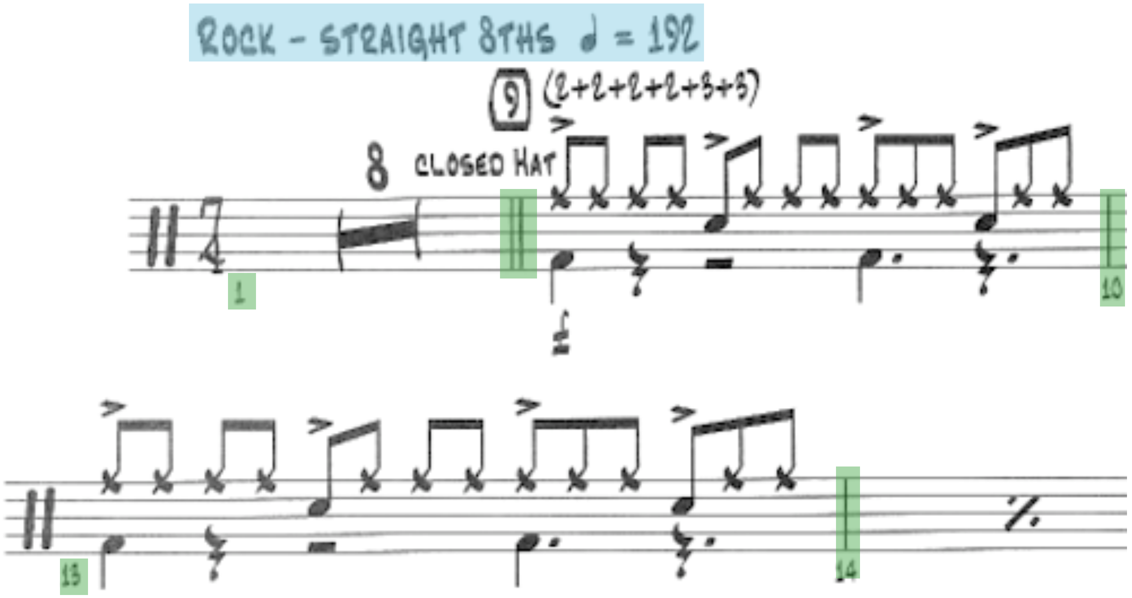
ISSUES OF CURRENT SYSTEMS

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



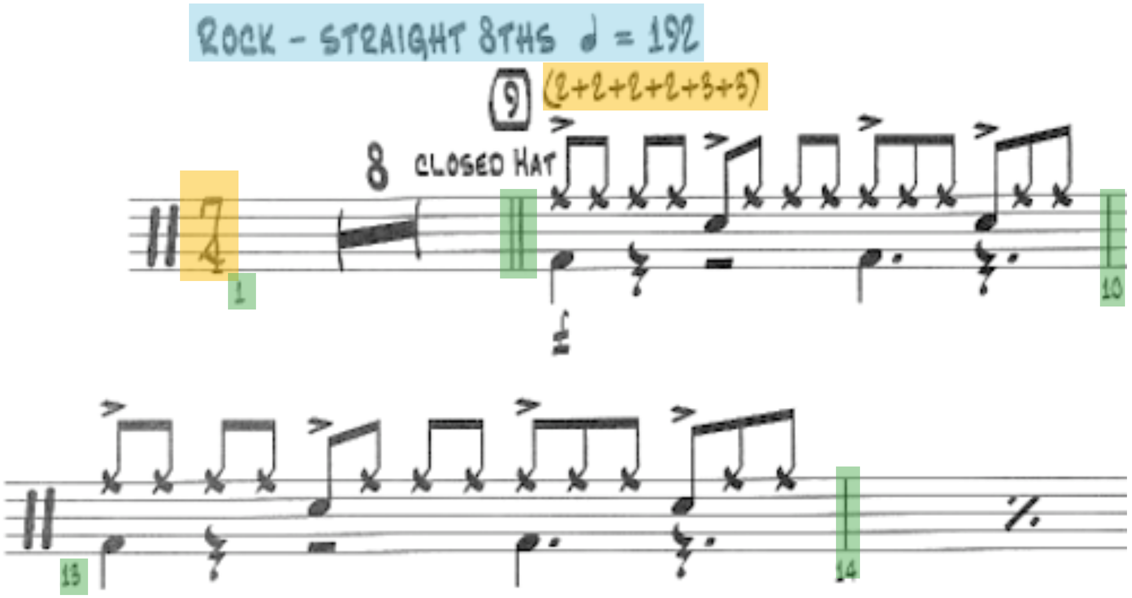
ISSUES OF CURRENT SYSTEMS

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



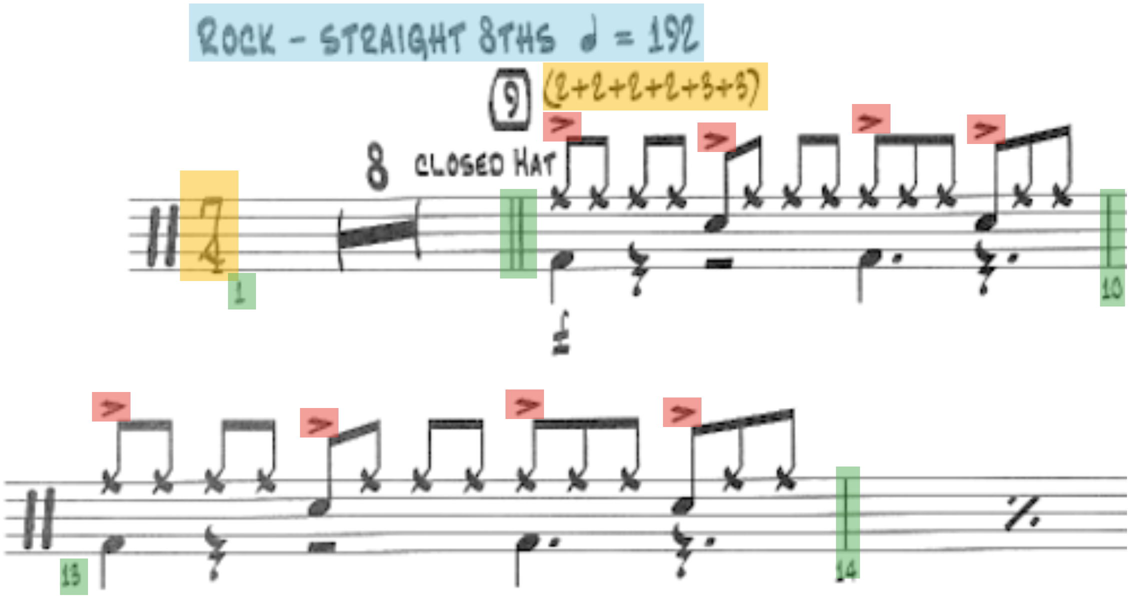
ISSUES OF CURRENT SYSTEMS

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



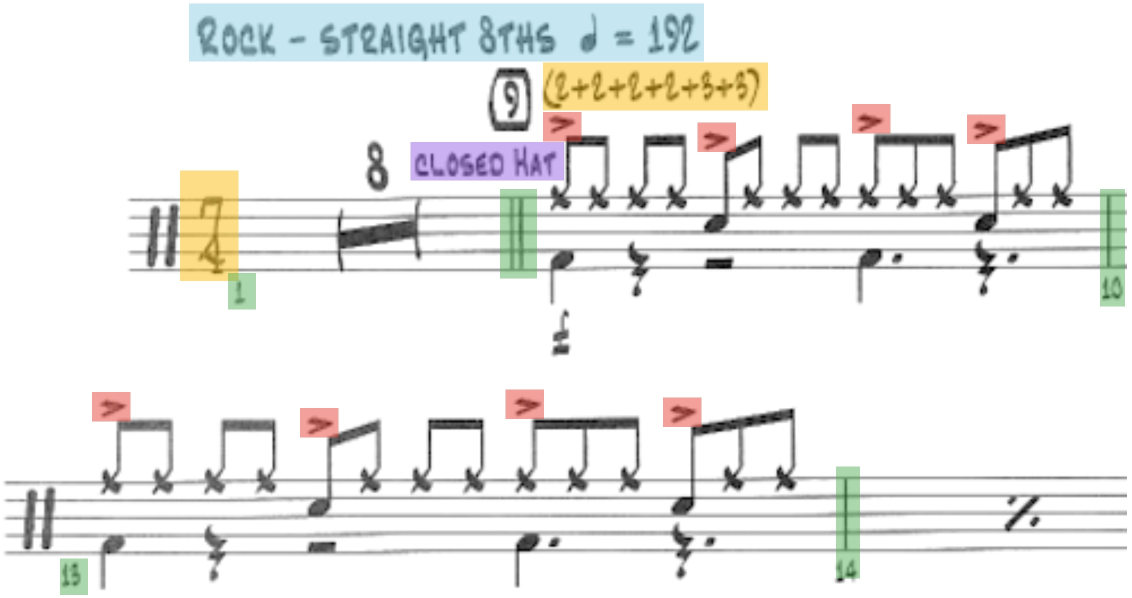
ISSUES OF CURRENT SYSTEMS

- 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



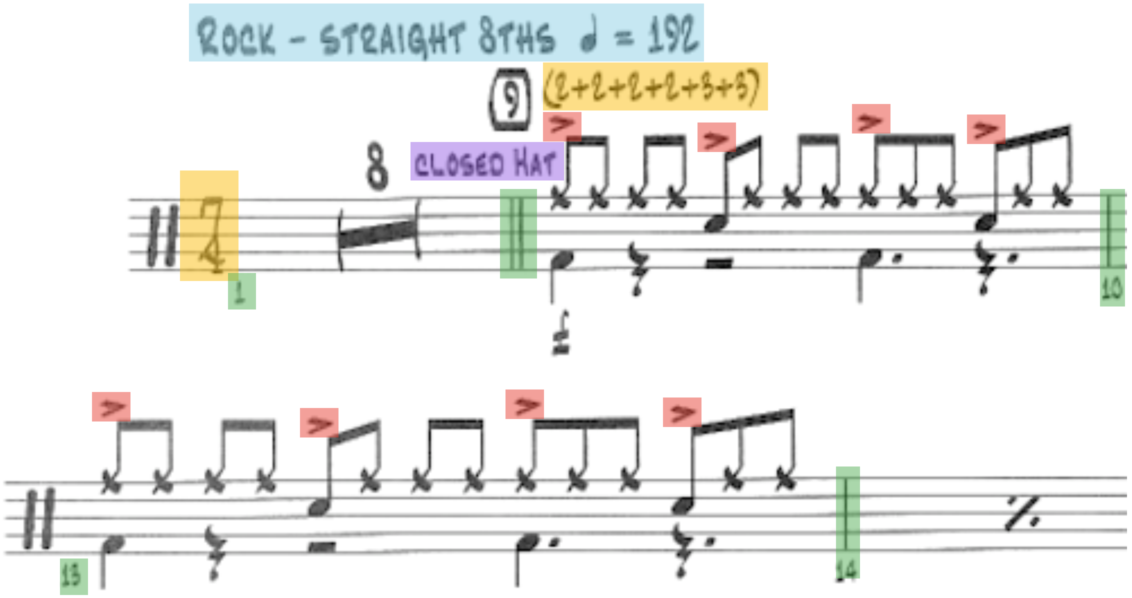
ISSUES OF CURRENT SYSTEMS

- 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



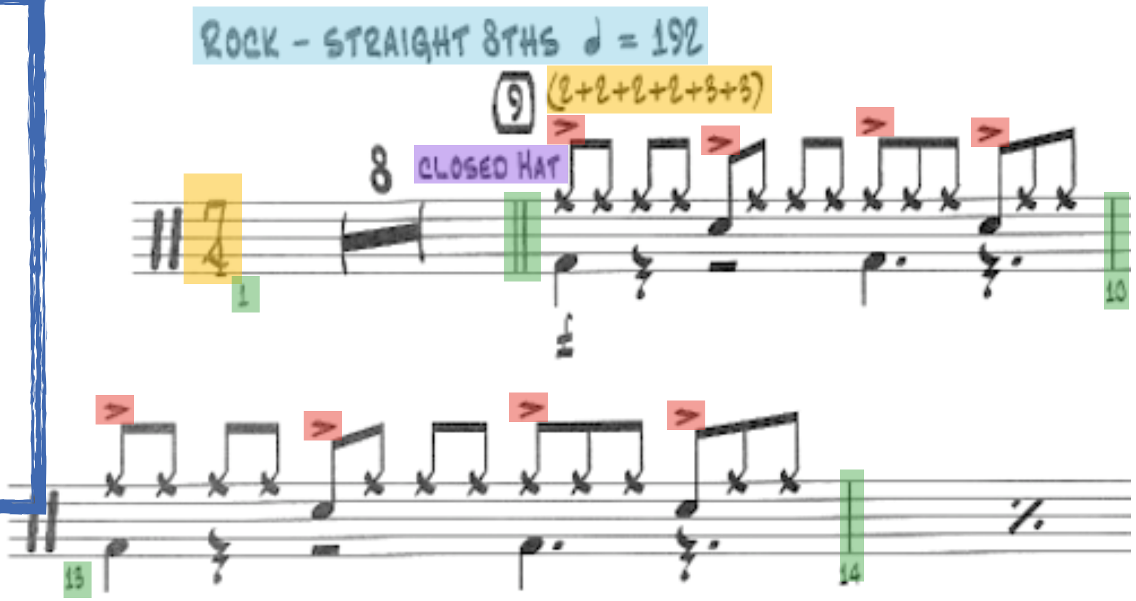
ISSUES OF CURRENT SYSTEMS

- 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.

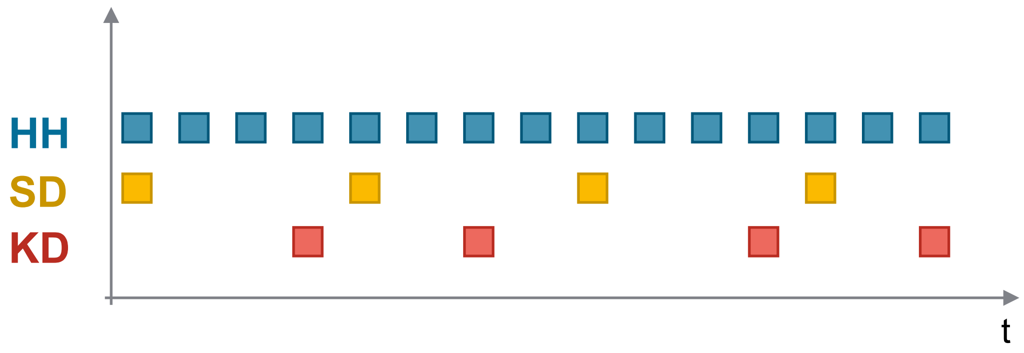


ISSUES OF CURRENT SYSTEMS

- 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.

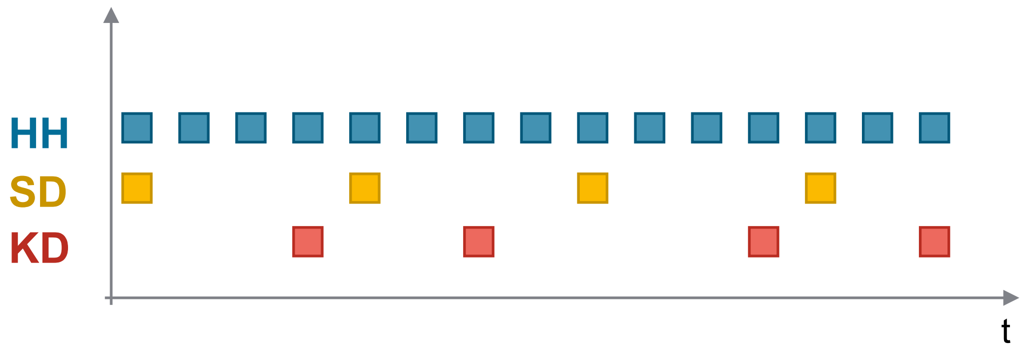


ADDITIONAL INFORMATION FOR TRANSCRIPTS



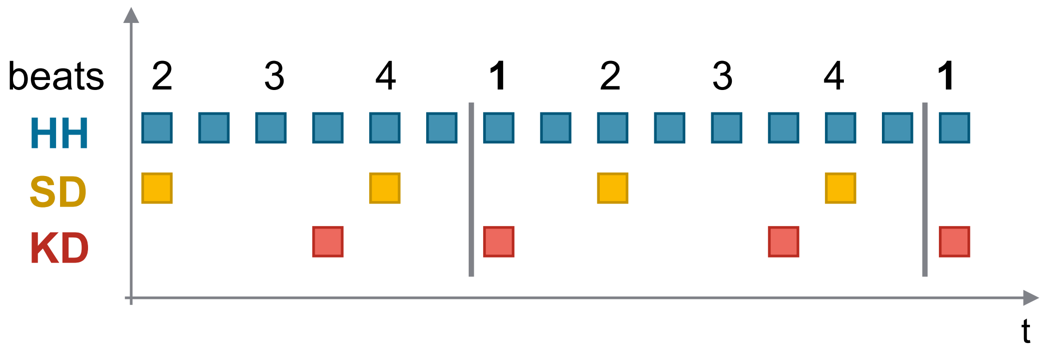
ADDITIONAL INFORMATION FOR TRANSCRIPTS

- Use **beat and downbeat tracking** to get:



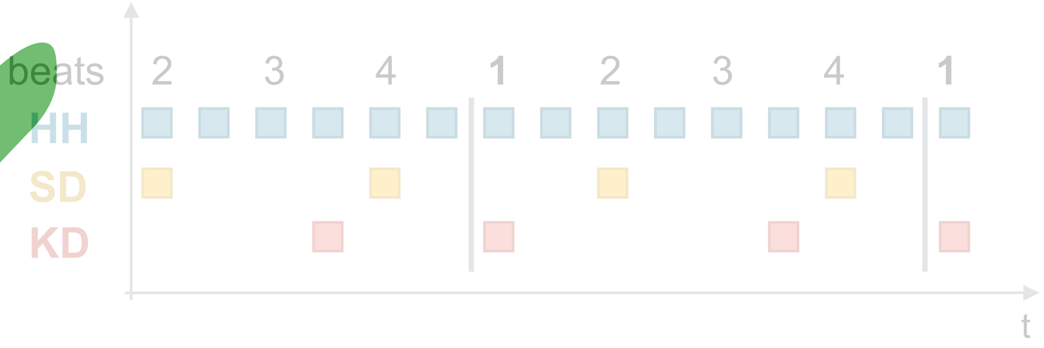
ADDITIONAL INFORMATION FOR TRANSCRIPTS

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



ADDITIONAL INFORMATION FOR TRANSCRIPTS

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

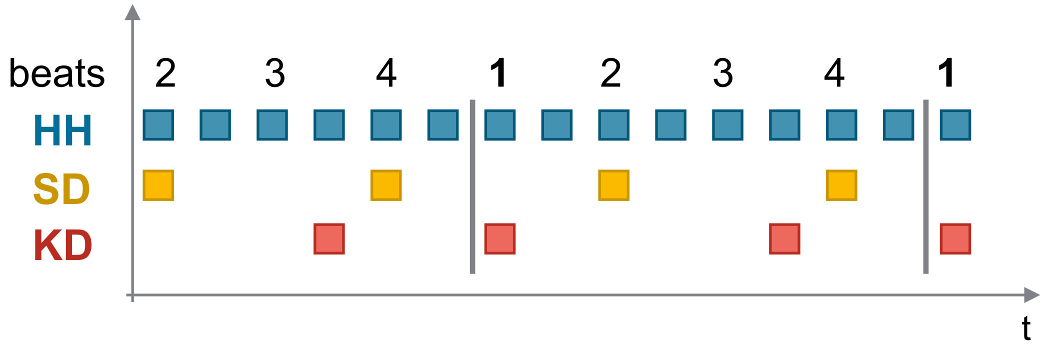


IMPROVE PERFORMANCE

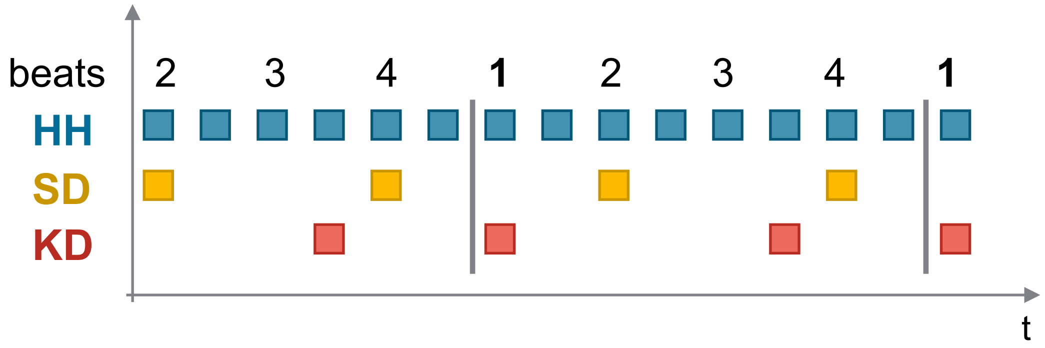
Three components to reach this goal:

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

1. LEVERAGE BEAT INFORMATION

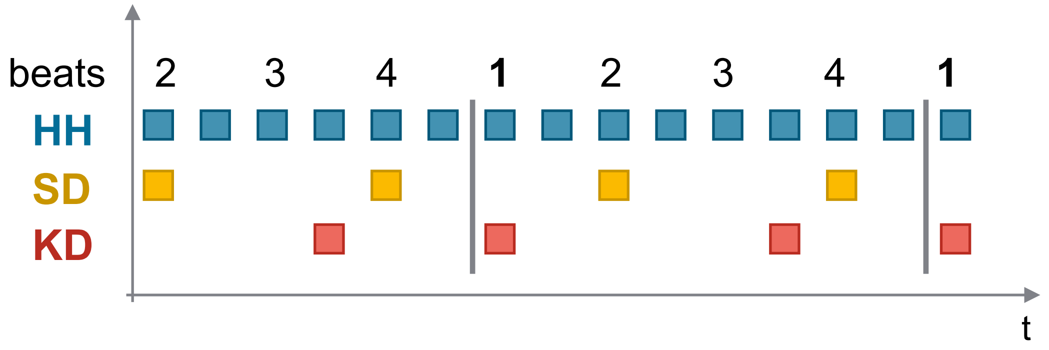


1. LEVERAGE BEAT INFORMATION



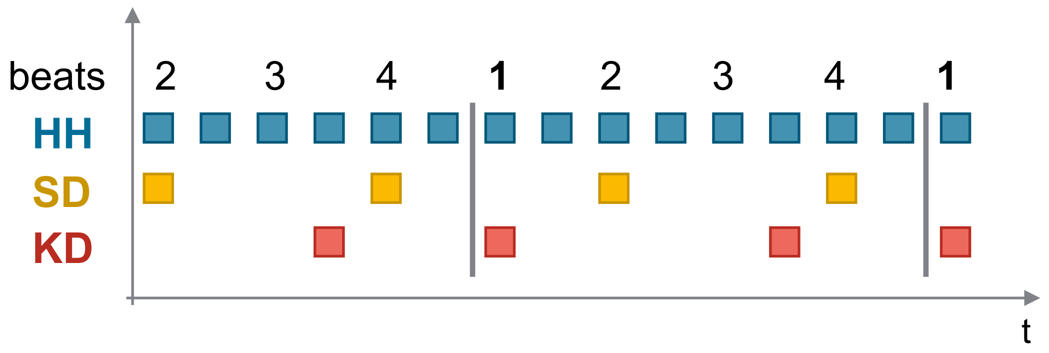
- Beats are **highly correlated** with drum patterns

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

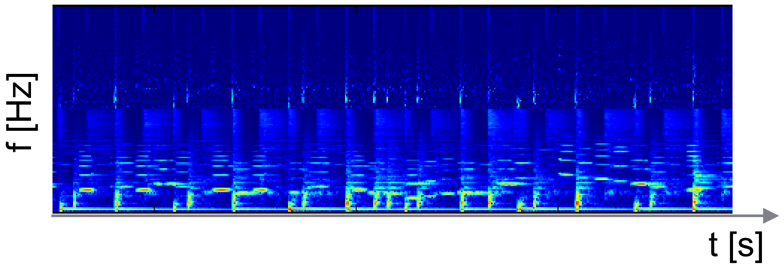


- Beats are **highly correlated** with drum patterns
- Assume that **prior knowledge** of beats is helpful for drum transcription (drum hit locations / repetitive patterns)
- Use **multi-task learning** for beats and drums

MULTI-TASK LEARNING

input

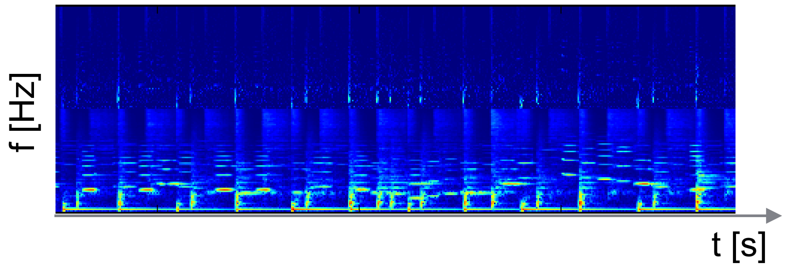
output



MULTI-TASK LEARNING

input

output

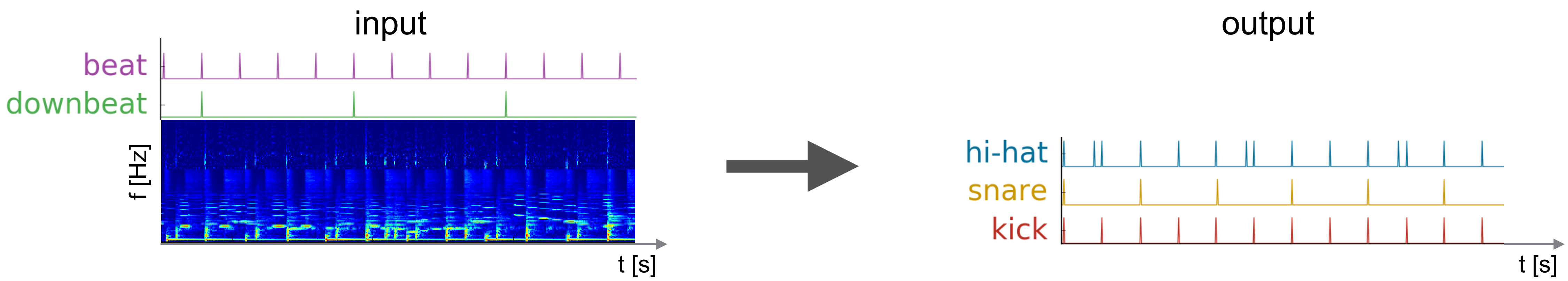


- Three experiments:

MULTI-TASK LEARNING

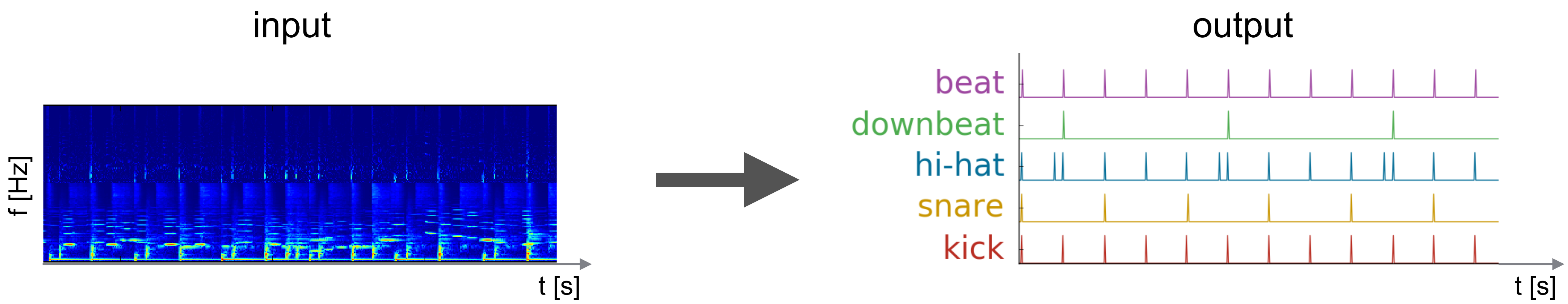
- Three experiments:
  - Training on drum targets (*DT*)

MULTI-TASK LEARNING



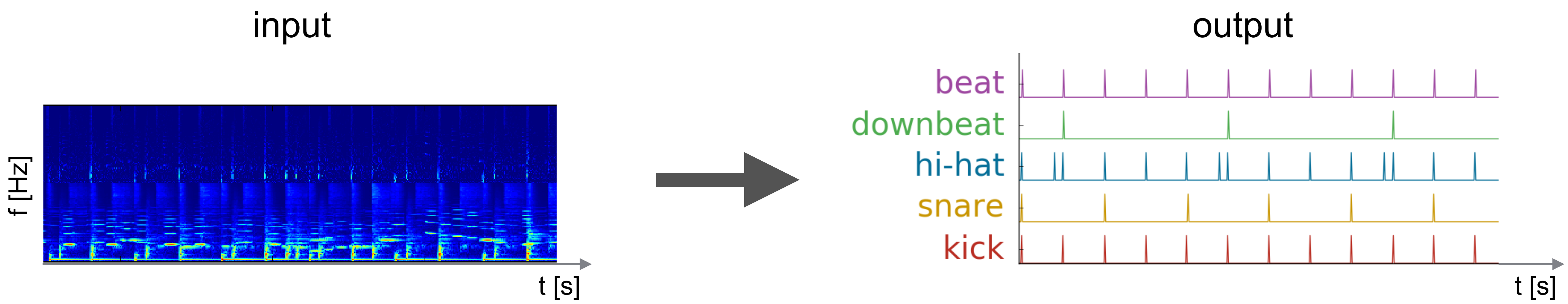
- Three experiments:
  - Training on drum targets (*DT*)
  - Training on drum targets with annotated beats as **additional input** features (*BF*)

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*)

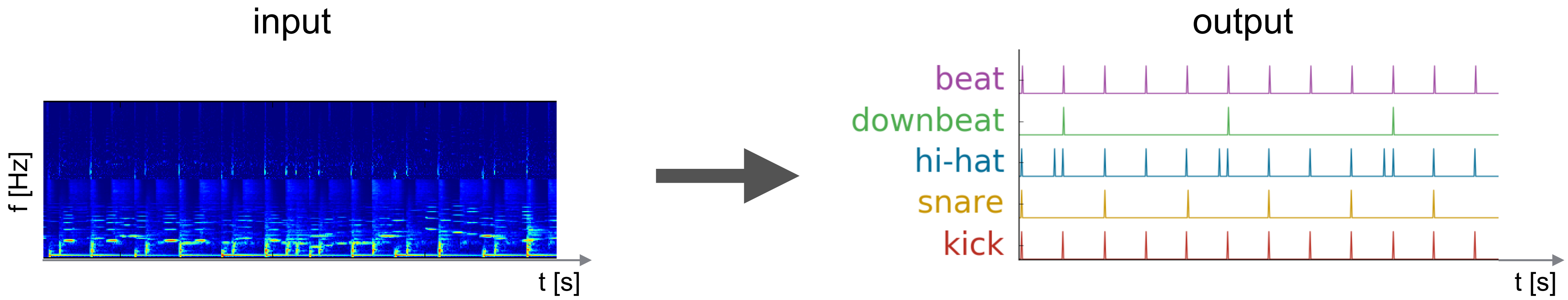
MULTI-TASK LEARNING



- 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*

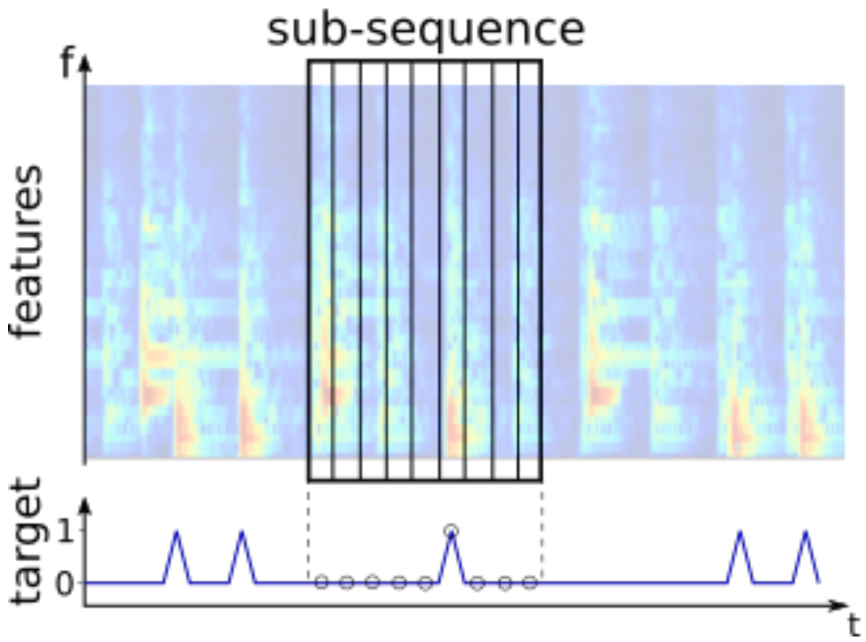
2. NETWORK MODELS — BASELINE MODELS

2. NETWORK MODELS — BASELINE MODELS

- Recurrent neural networks

2. NETWORK MODELS — BASELINE MODELS

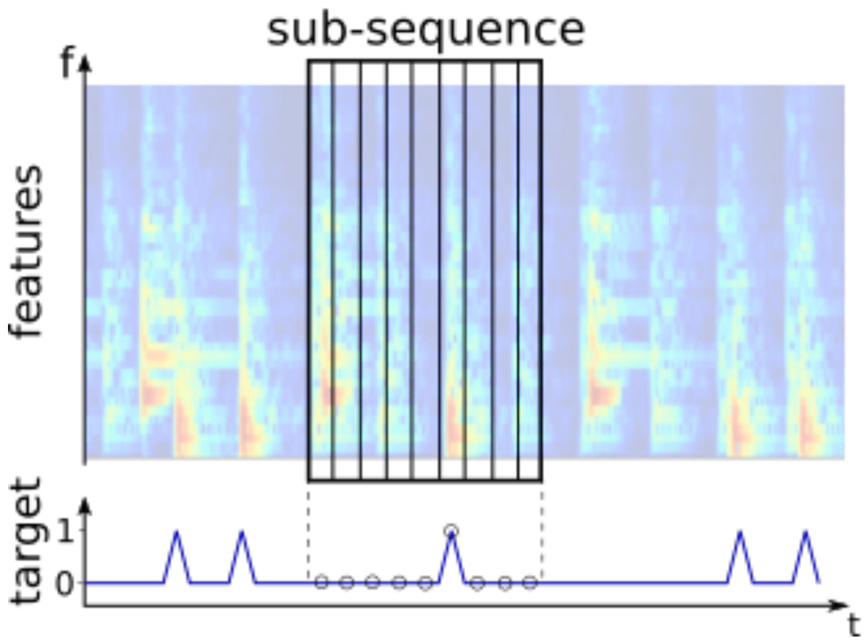
- Recurrent neural networks
  - Recurrent connections act as **memory**
  - Processing of **sequential data**



RNN train data sample

2. NETWORK MODELS — BASELINE MODELS

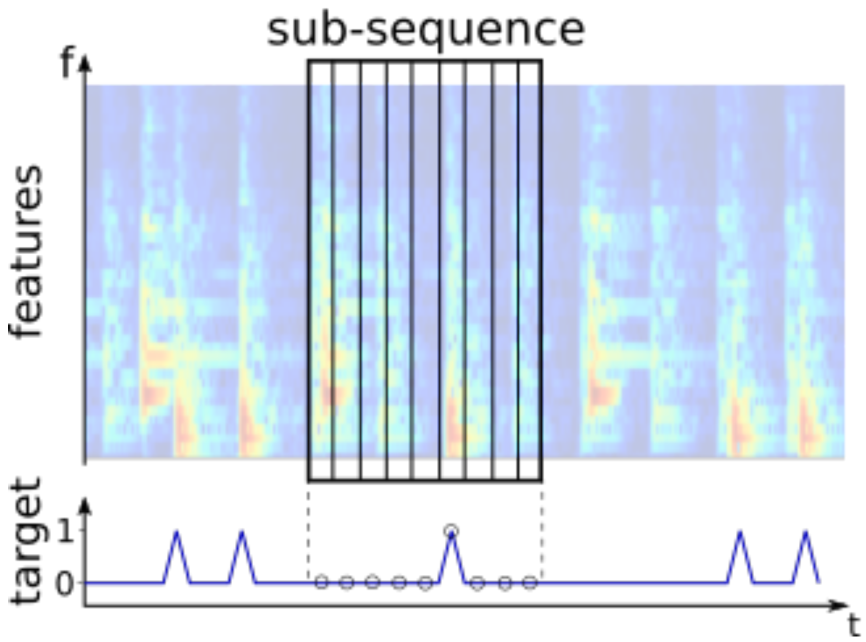
- Recurrent neural networks
  - Recurrent connections act as **memory**
  - Processing of **sequential data**
  - Work well for drum detection and beat tracking
    [Böck et al. ISMIR'16]



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 [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

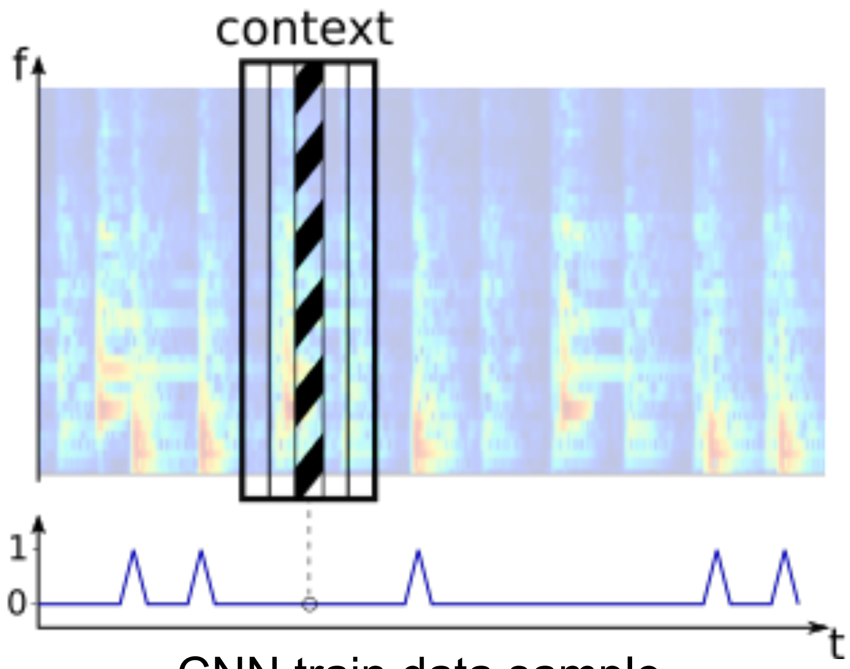


RNN train data sample

2. NETWORK MODELS — NEW FOR DT

2. NETWORK MODELS — NEW FOR DT

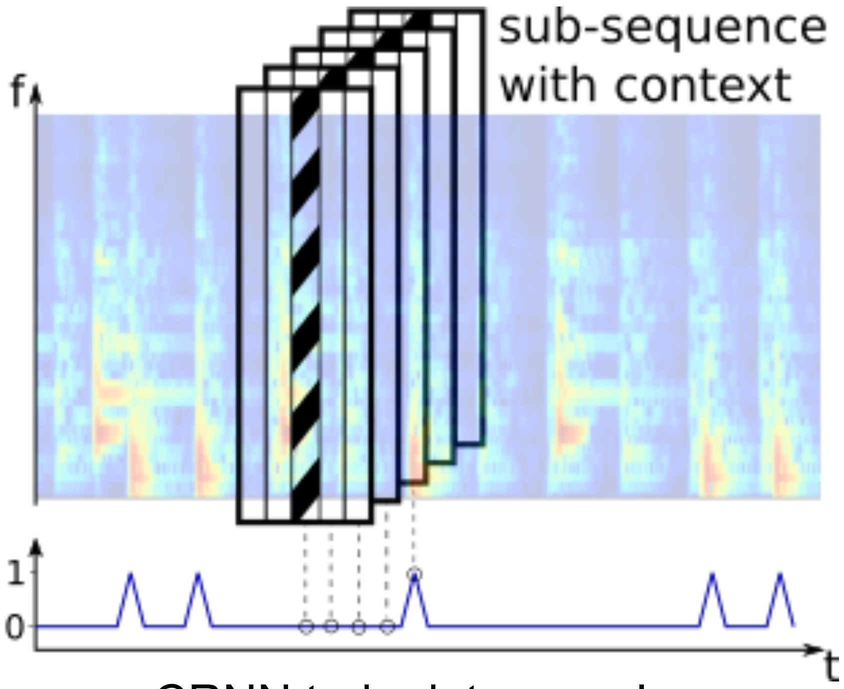
- Convolutional NN (**CNN**)
  - Convolutions capture **local correlations**
  - **Acoustic modeling** of drum sounds



CNN train data sample

2. NETWORK MODELS — NEW FOR DT

- Convolutional NN (**CNN**)
  - Convolutions capture **local correlations**
  - **Acoustic modeling** of drum sounds
- Convolutional BDRNN (**CRNN**)
  - ***"best of both worlds"***
  - Low-level CNN for **acoustic modeling**
  - Higher-level RNN for **repetitive pattern modeling**



CRNN train data sample

NETWORK MODELS

	Frames	Context	Conv. Layers	Rec. Layers	Dense Layers
BDRNN (S)	100	—	—	2x50 GRU	—
BDRNN (L)	400	—	—	3x30 GRU	—
CNN (S)	—	9	2 x 32 3x3 filt. 3x3 max pooling 2 x 64 3x3 filt. 3x3 max pooling all w/ batch norm.	—	2x256
CNN (L)	—	25		—	2x256
CRNN (S)	100	9		2x50 GRU	—
CRNN (L)	400	13		3x60 GRU	—
tsRNN	*state-of-the-art baseline [Vogl et al. ICASSP'17]*				

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

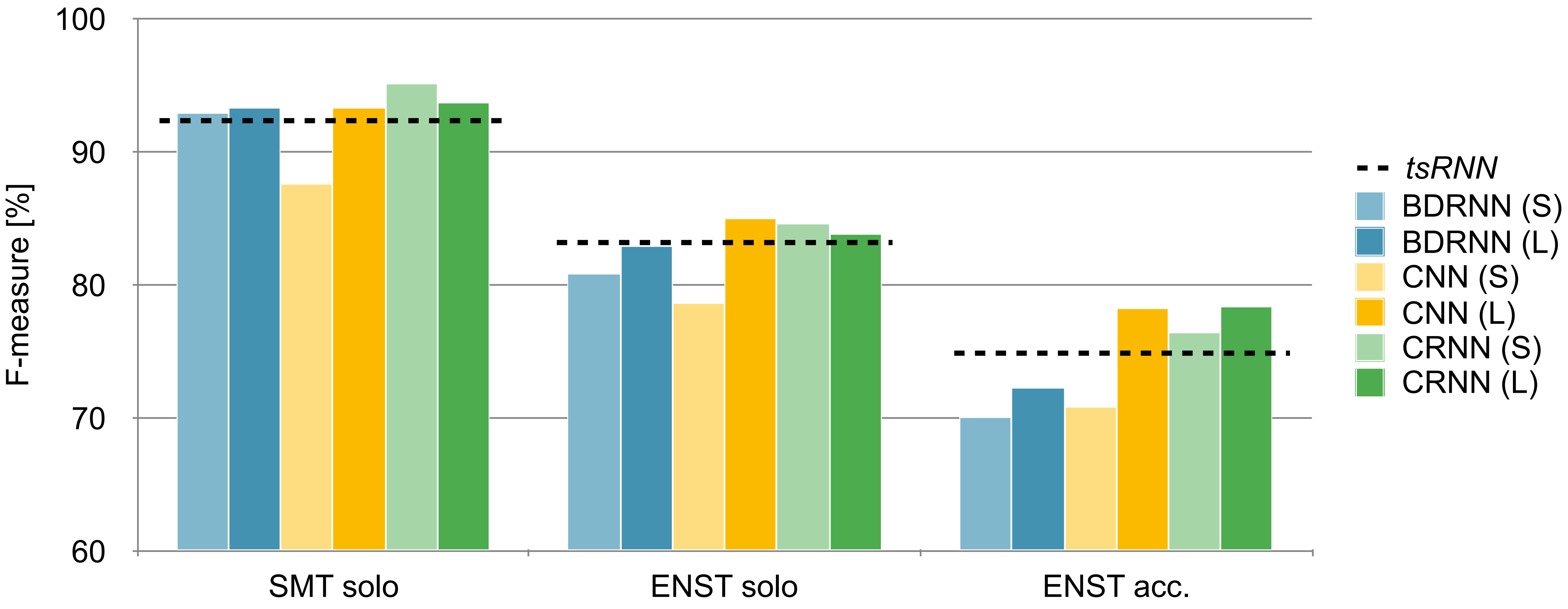
- **ENST-Drums** [Gillet and Richard 2006]
  - Recordings, three drummers on different drum kits, **optional accompaniment**
  - 64 tracks, total: **1h**, onsets: **22391** + training samples

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

DT 3-FOLD CV RESULTS ON CLASSIC DATASETS



3. 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

♫ ♫

3. 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

3. 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

♫ ♫

3. NEW DATASETS (DRUMS AND BEATS)

NEW!

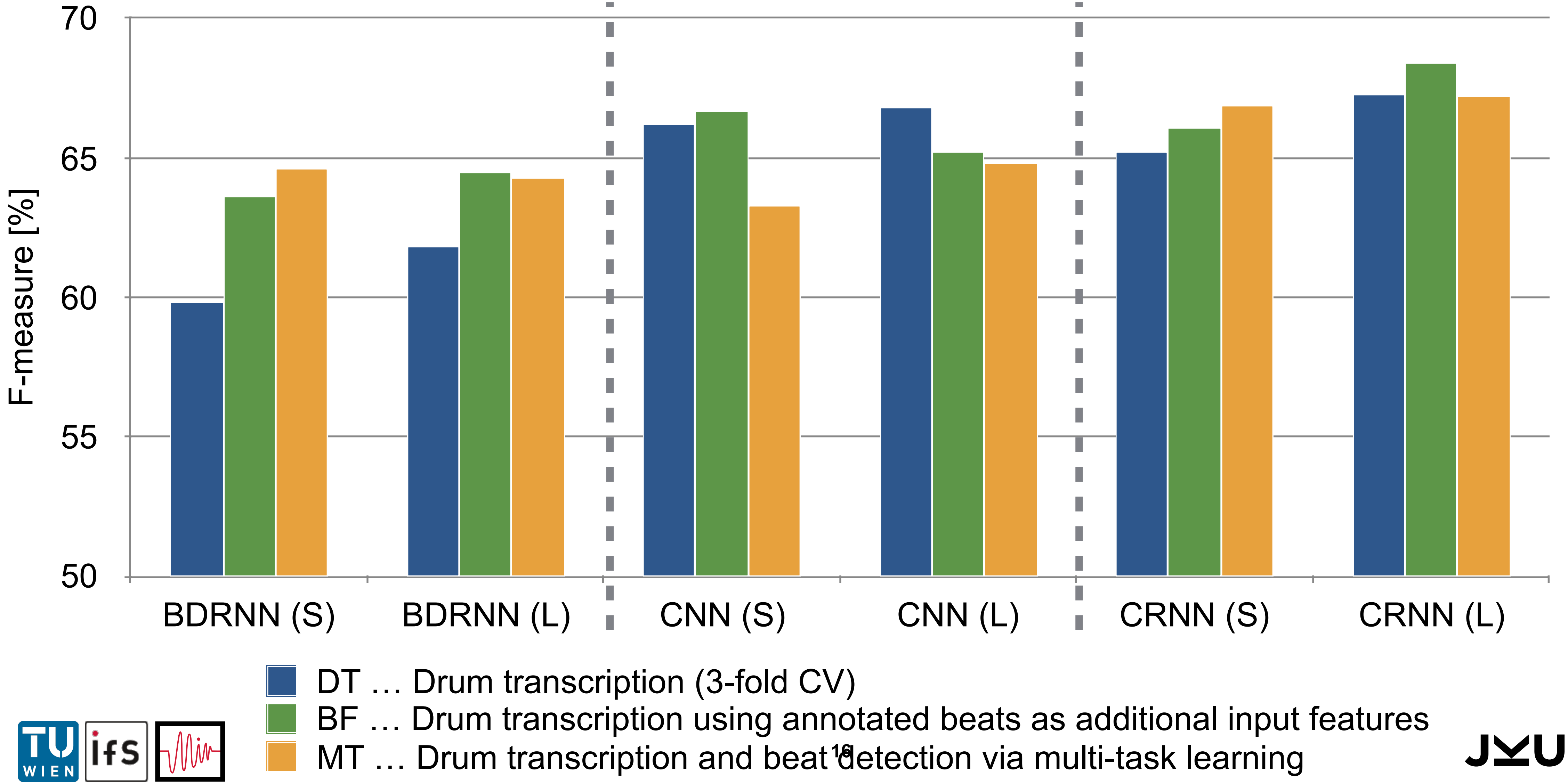
RBMA13-Drums [http://ifs.tuwien.ac.at/~vogl/datasets/]

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

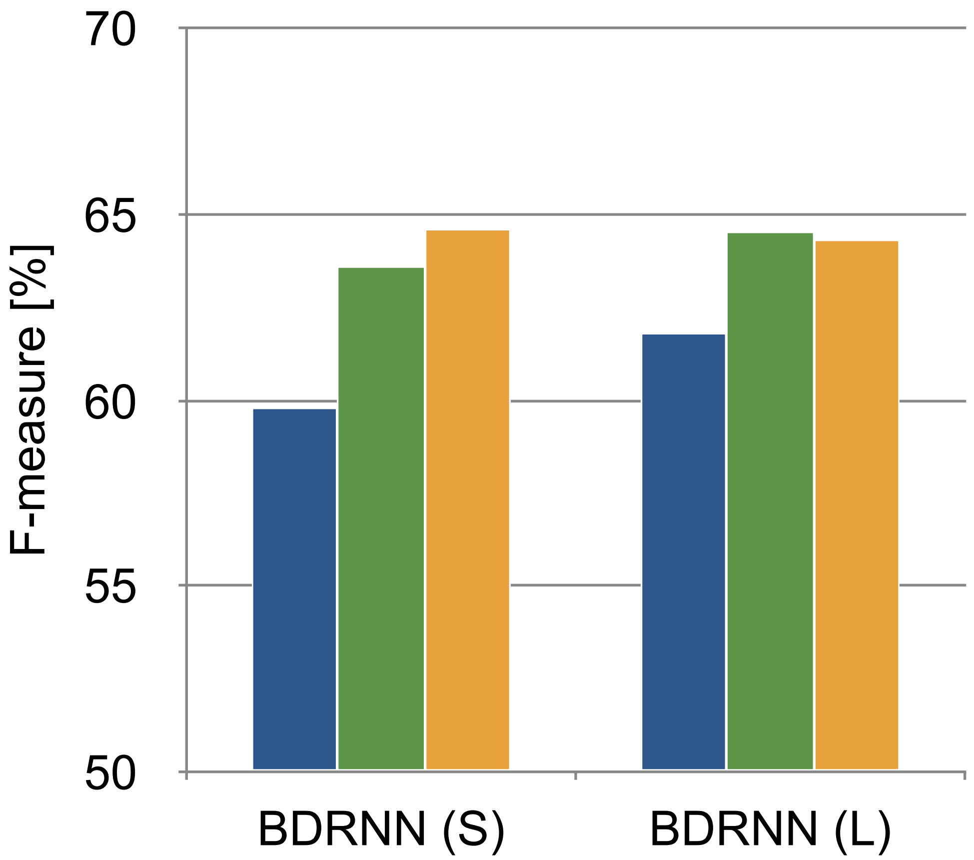
♫ ♫

- **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



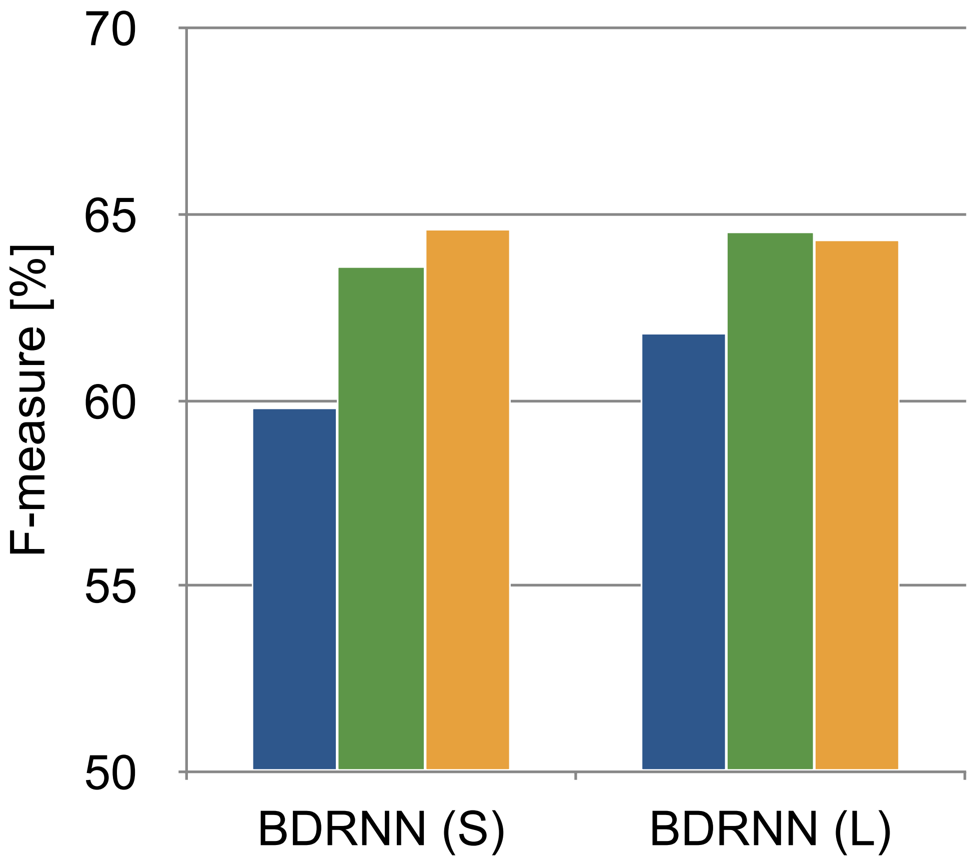
RESULTS ON RBMA13: BDRNNs



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: BDRNNs

Impact on bi-directional RNNs:

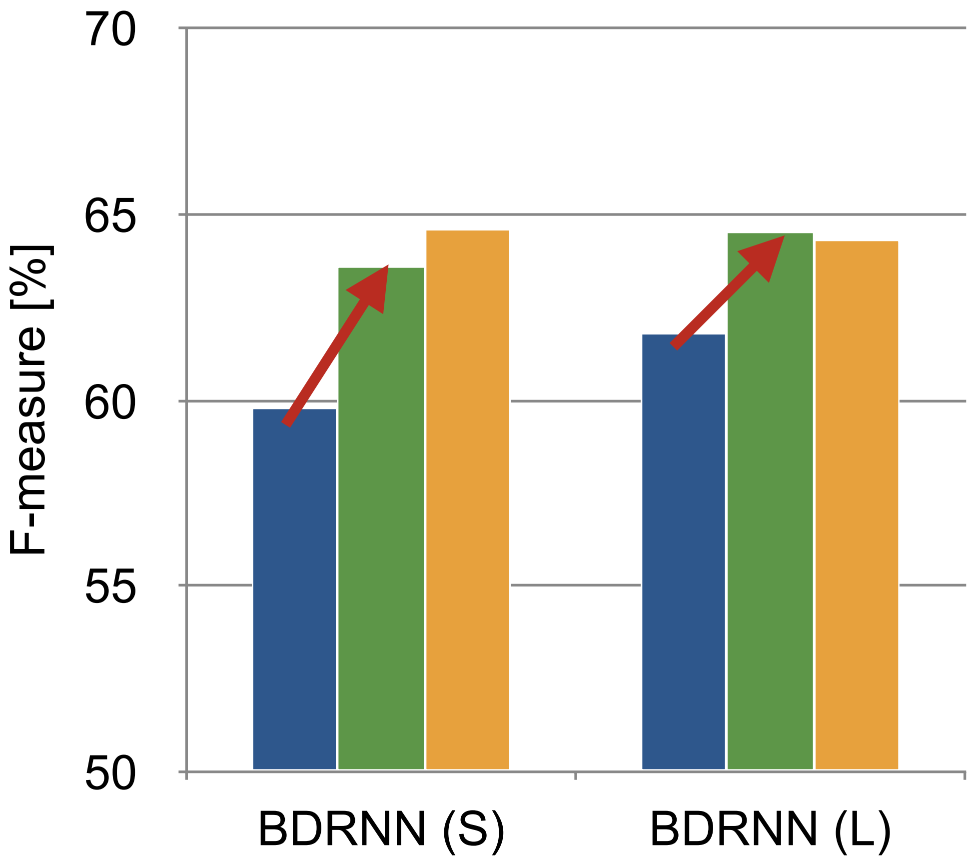


DT … Drum transcription (3-fold CV)
BF … Drum transcription using annotated beats as additional input features
MT … Drum transcription and beat detection via multi-task learning

RESULTS ON RBMA13: BDRNNs

Impact on bi-directional RNNs:

- BF improves for both models ✔

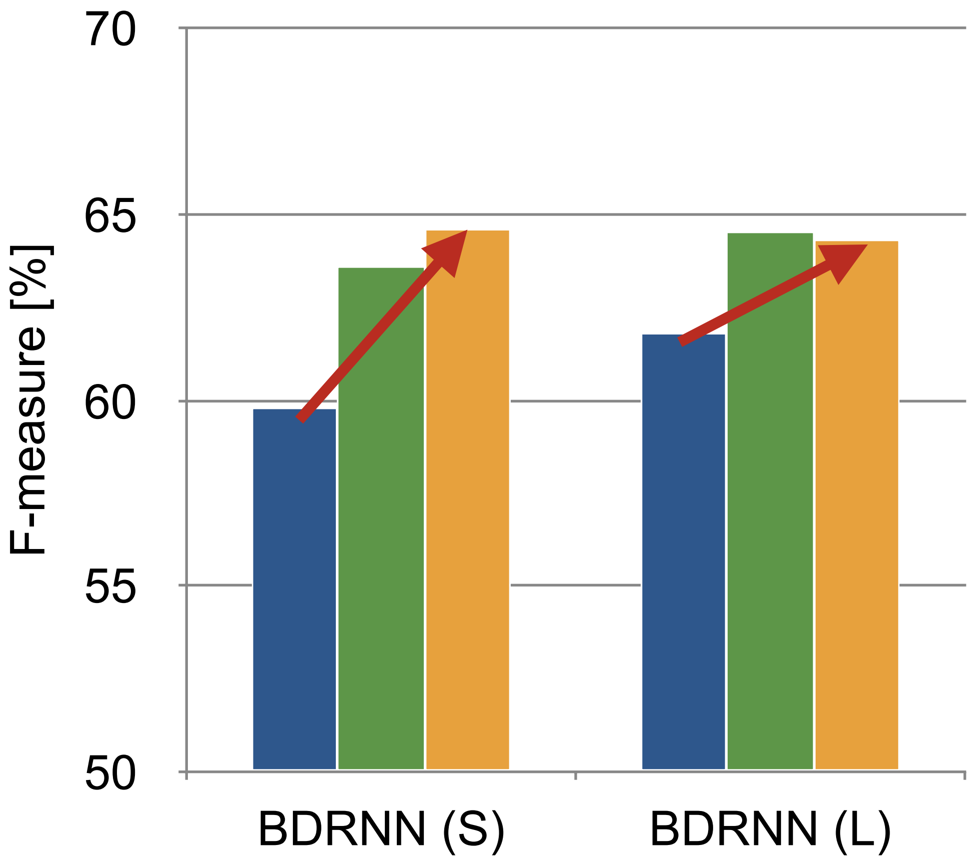


DT … Drum transcription (3-fold CV)
BF … Drum transcription using annotated beats as additional input features
MT … Drum transcription and beat detection via multi-task learning

RESULTS ON RBMA13: BDRNNs

Impact on bi-directional RNNs:

- BF improves for both models ✔
- MT improves for both models ✔

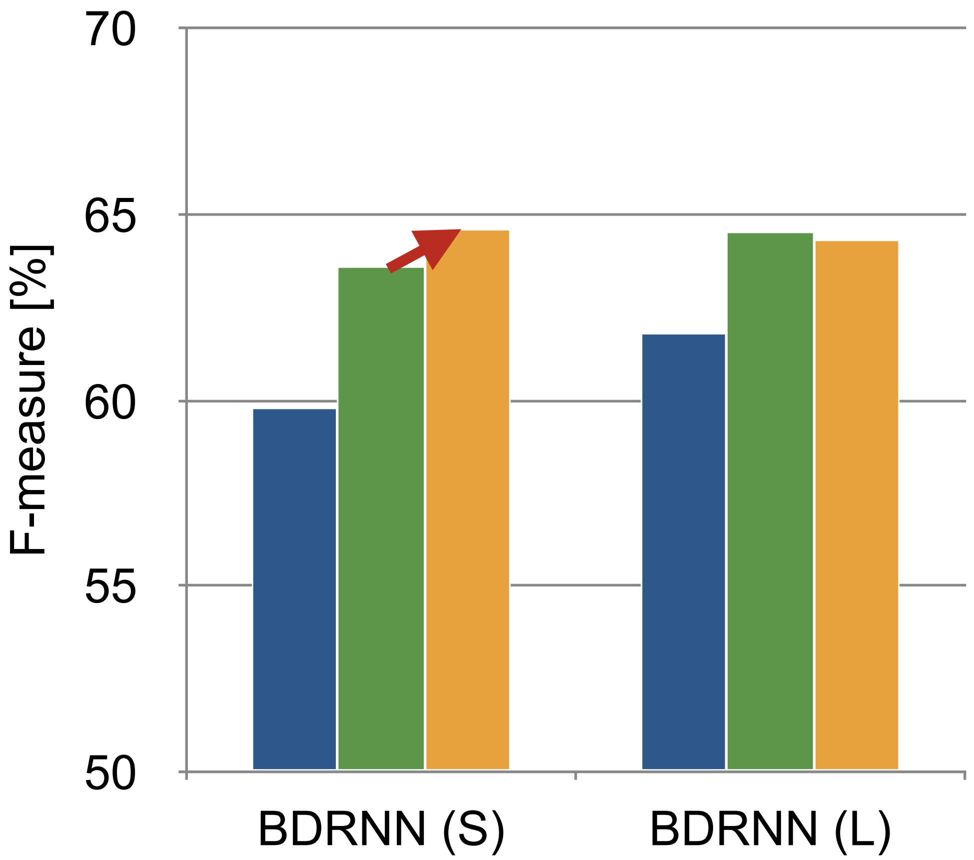


- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning

RESULTS ON RBMA13: BDRNNs

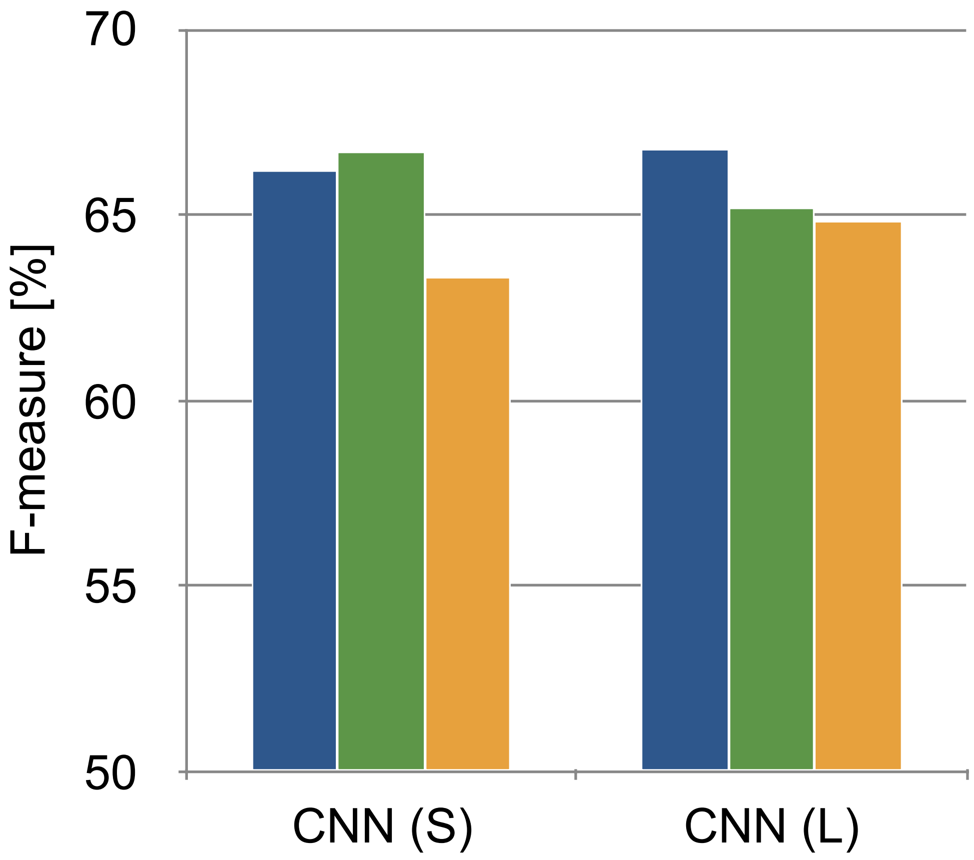
Impact on bi-directional RNNs:

- BF improves for both models ✔
- MT improves for both models ✔
- MT even better than BF for small 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

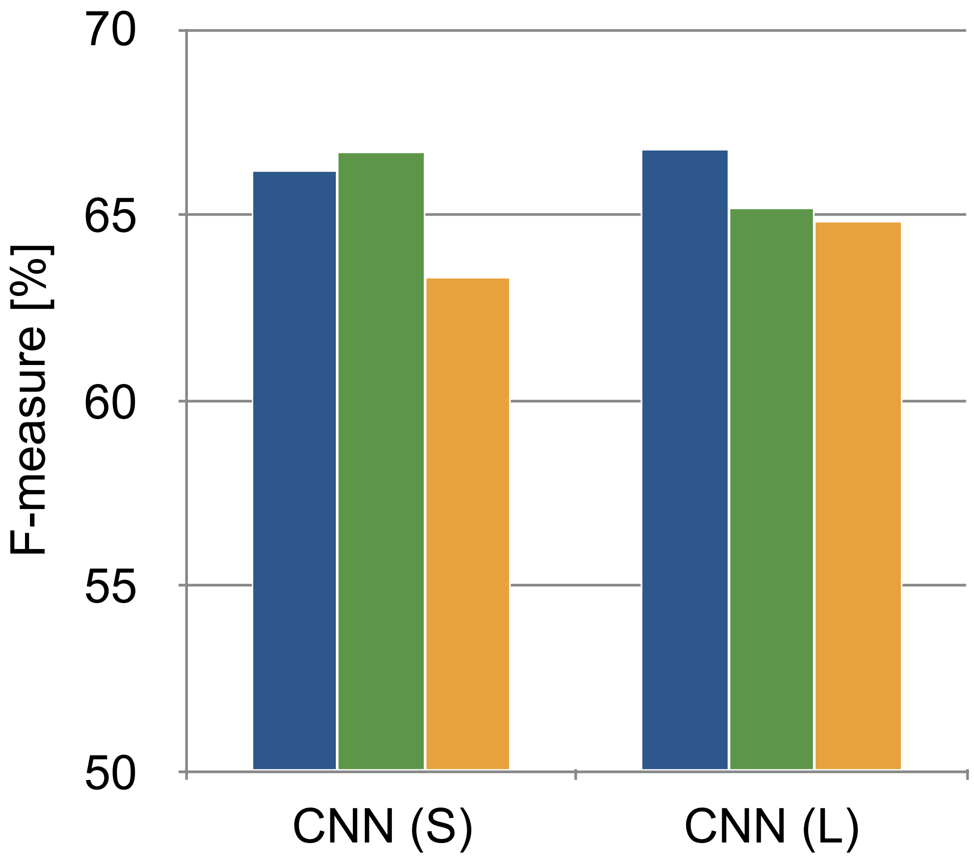
RESULTS ON RBMA13: CNNs



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:

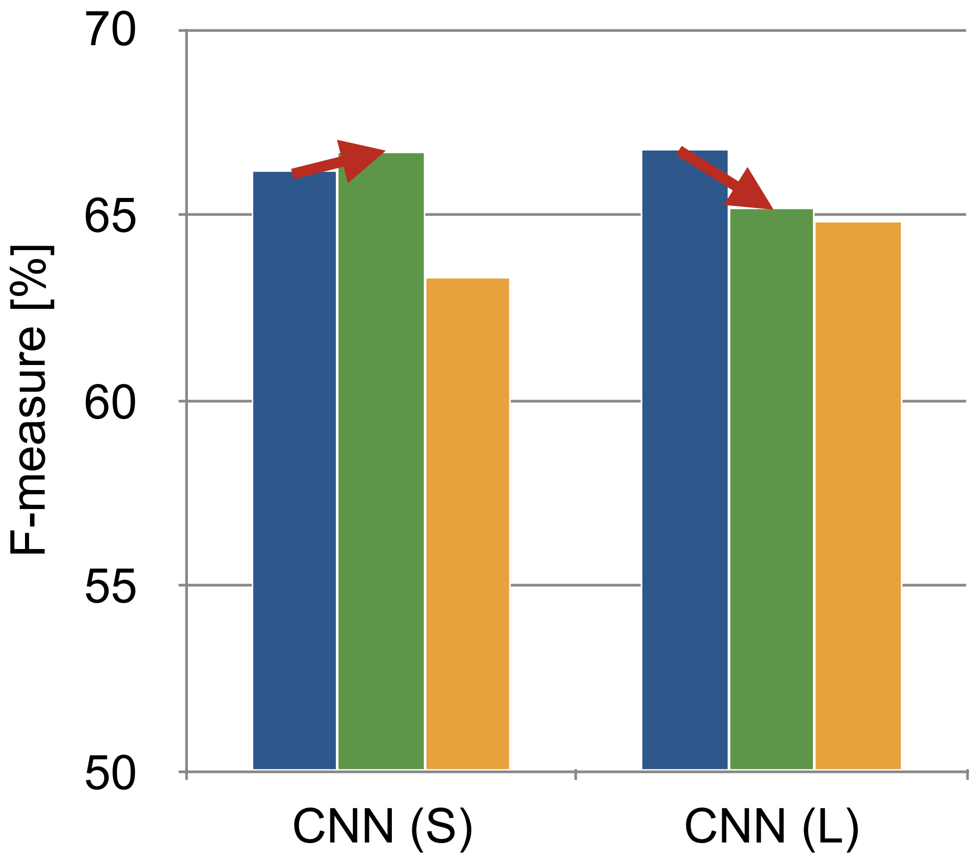


DT … Drum transcription (3-fold CV)
BF … Drum transcription using annotated beats as additional input features
MT … Drum transcription and beat detection via multi-task learning

RESULTS ON RBMA13: CNNs

Impact on CNNs:

- BF inconsistent

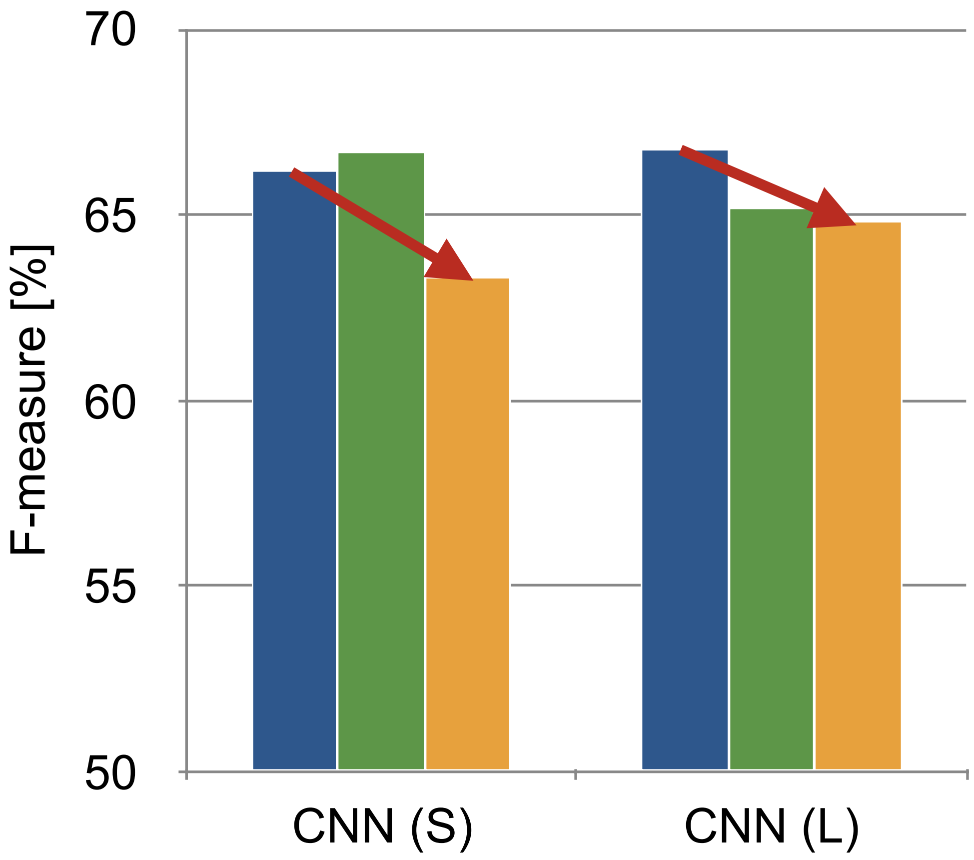


DT … Drum transcription (3-fold CV)
BF … Drum transcription using annotated beats as additional input features
MT … Drum transcription and beat detection via multi-task learning

RESULTS ON RBMA13: CNNs

Impact on CNNs:

- BF inconsistent
- MT declines for both models

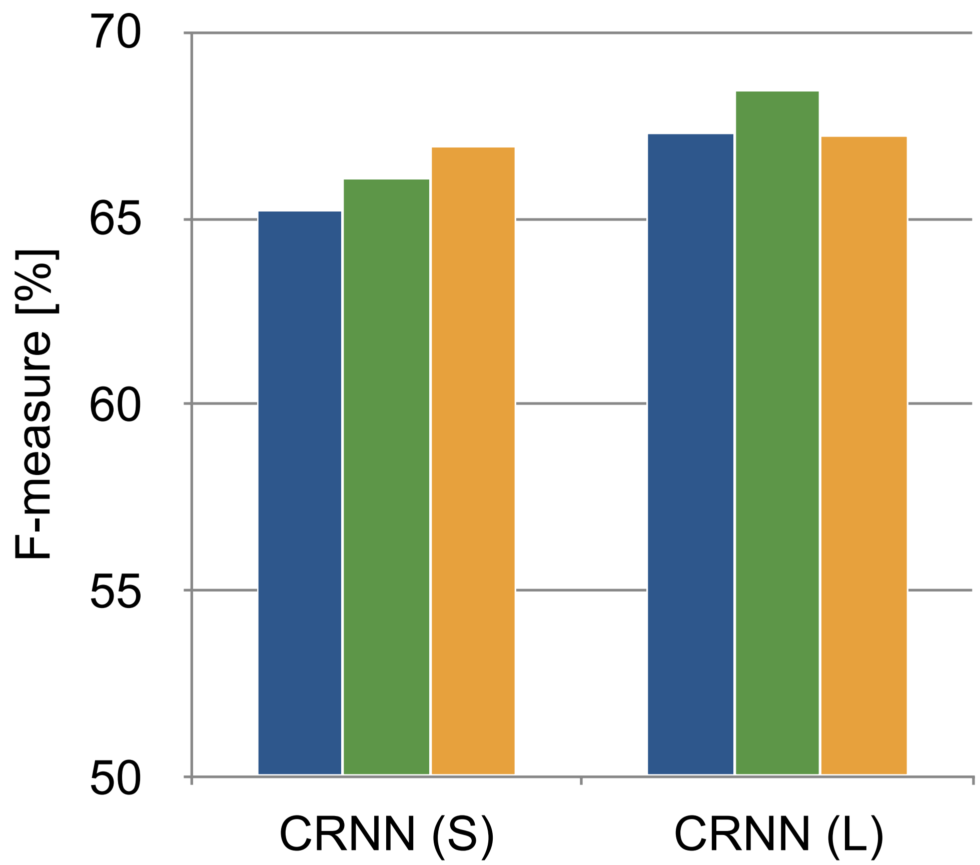


DT … Drum transcription (3-fold CV)

BF … Drum transcription using annotated beats as additional input features

MT … Drum transcription and beat detection via multi-task learning

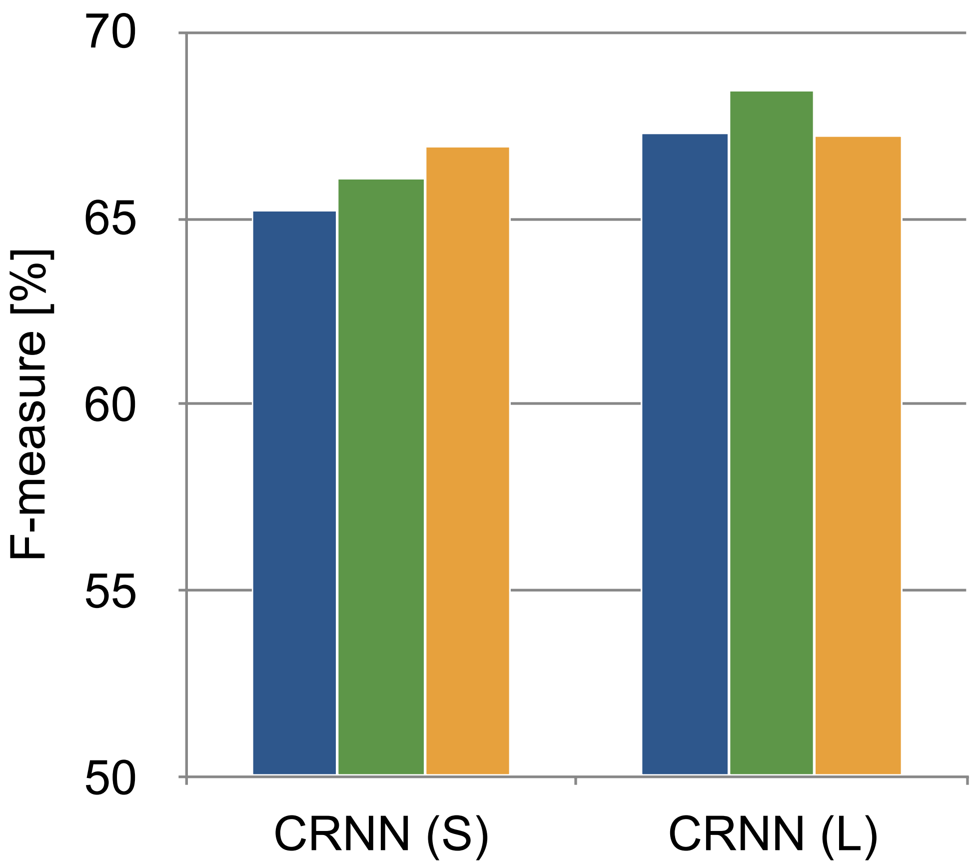
RESULTS ON RBMA13: CRNNs



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:

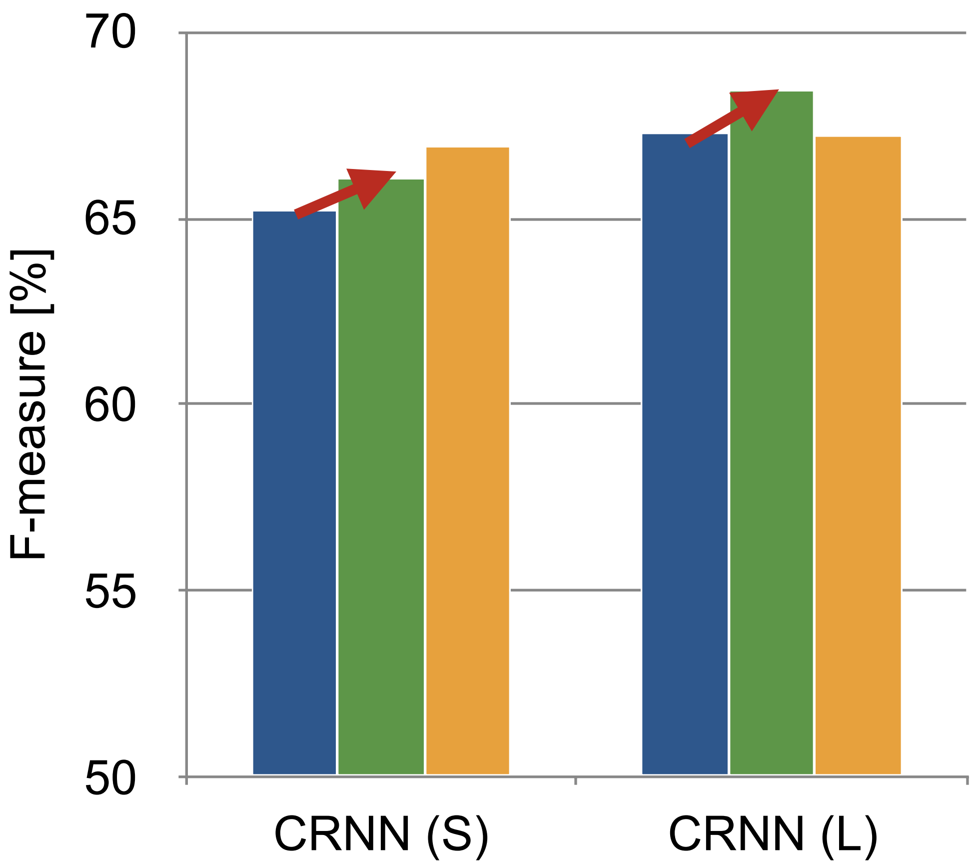


DT … Drum transcription (3-fold CV)
BF … Drum transcription using annotated beats as additional input features
MT … Drum transcription and beat detection via multi-task learning

RESULTS ON RBMA13: CRNNs

Impact on CRNNs:

- BF improves for both models ✔

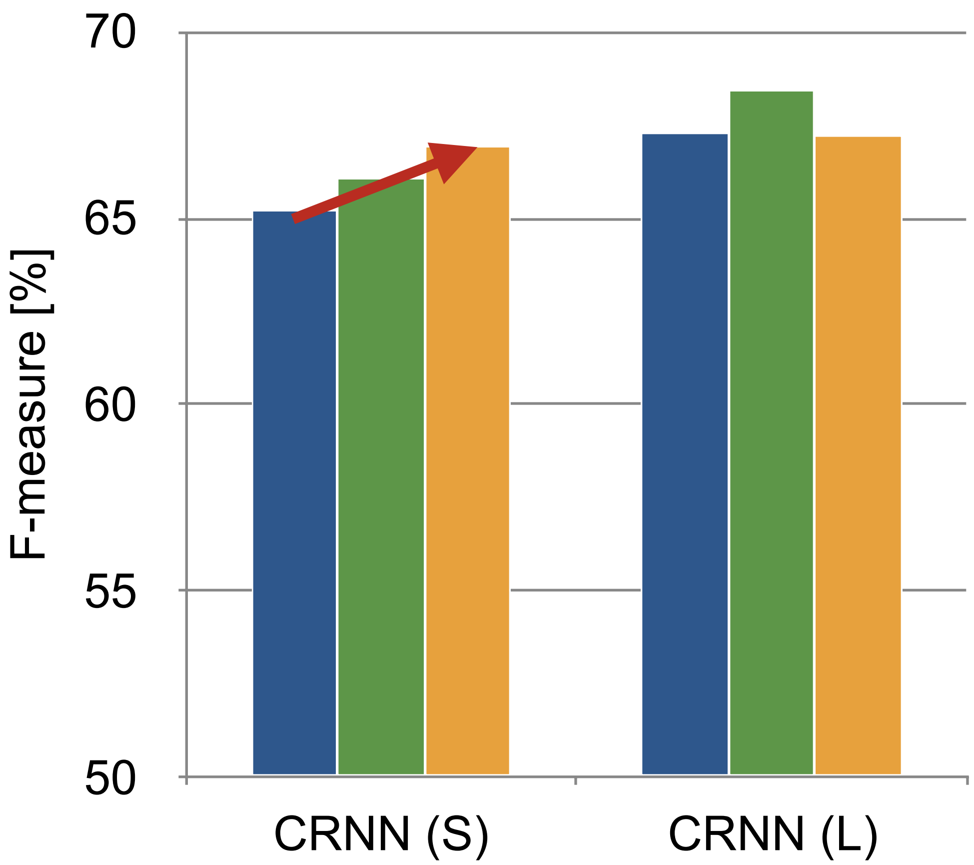


- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning

RESULTS ON RBMA13: CRNNs

Impact on CRNNs:

- BF improves for both models ✔
- MT improves for small models ✔

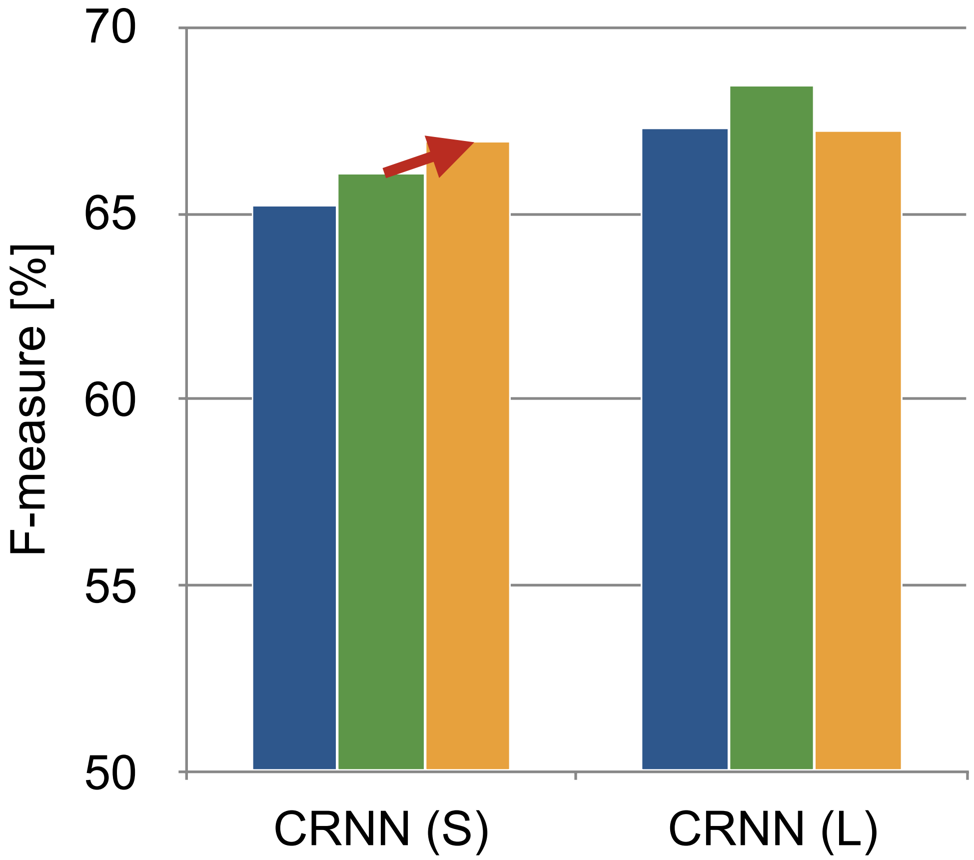


- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning

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 !

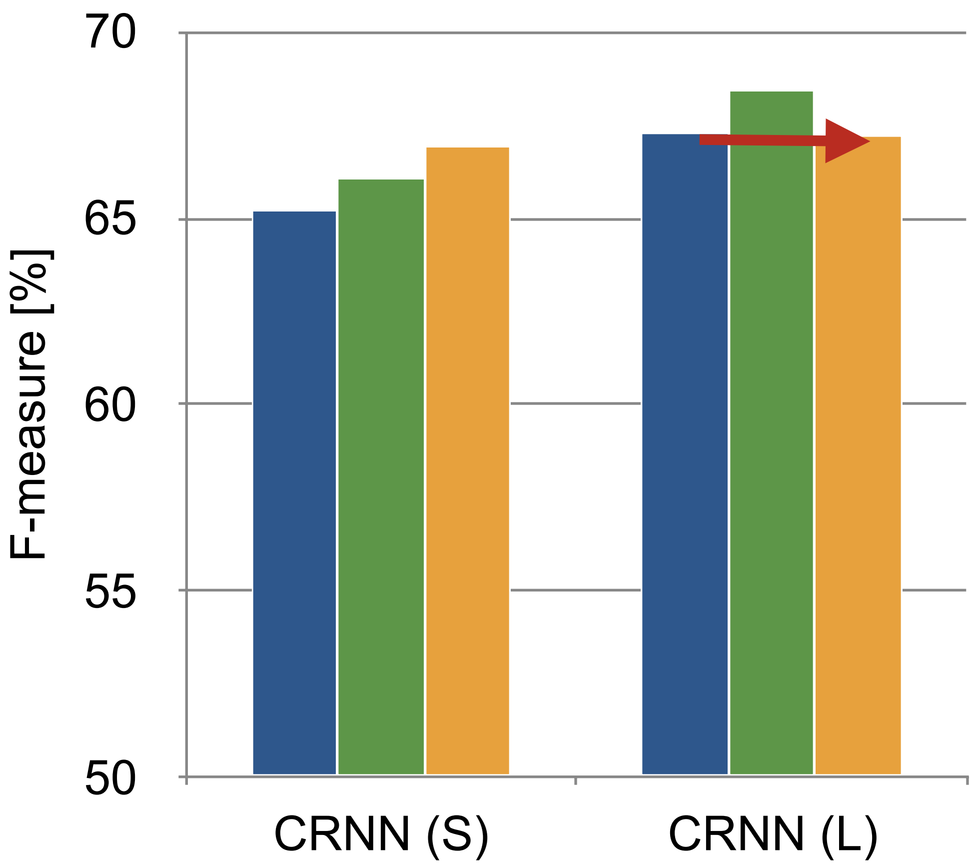


DT … Drum transcription (3-fold CV)
BF … Drum transcription using annotated beats as additional input features
MT … Drum transcription and beat detection via multi-task learning

RESULTS ON RBMA13: CRNNs

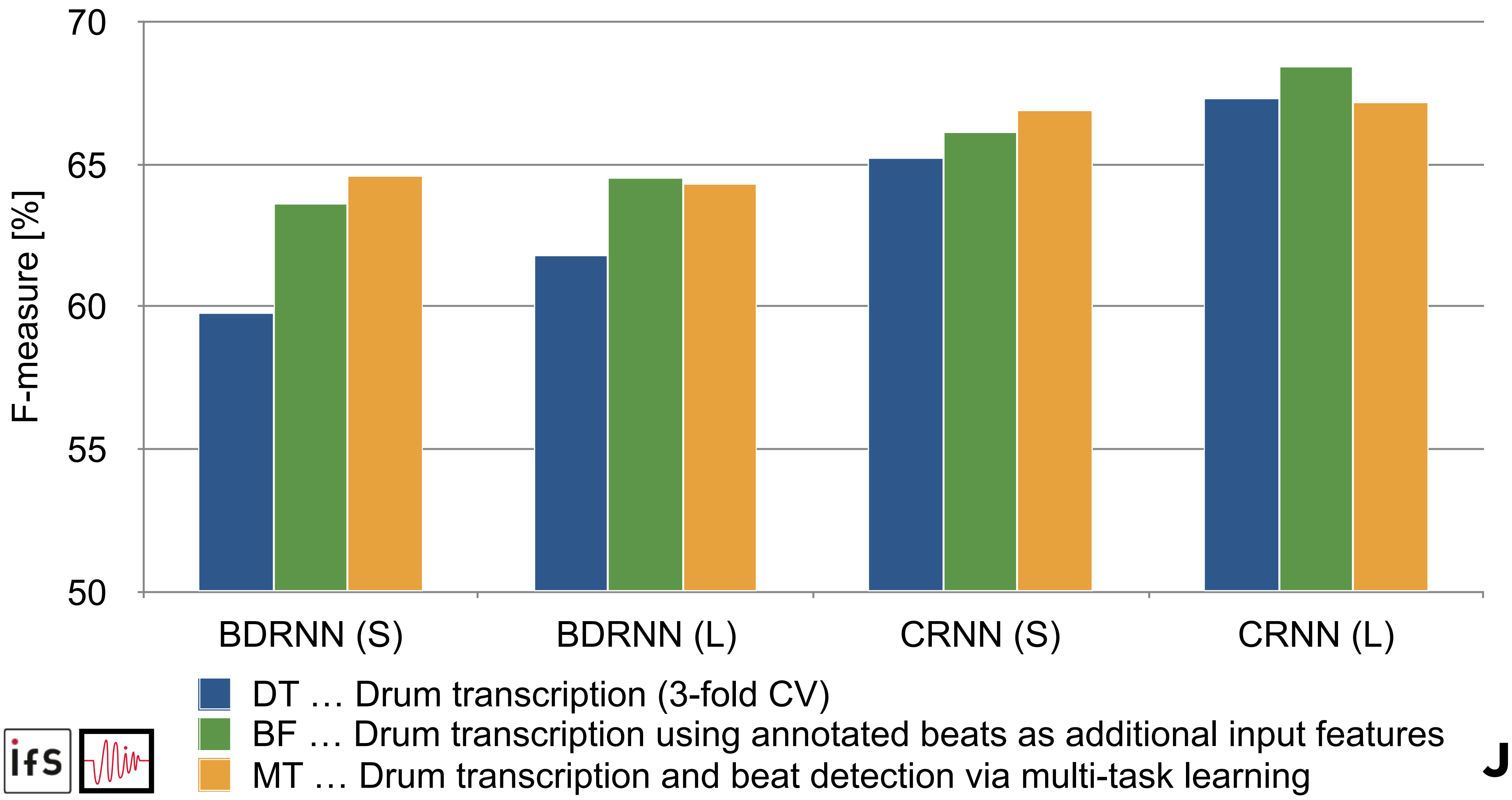
Impact on CRNNs:

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

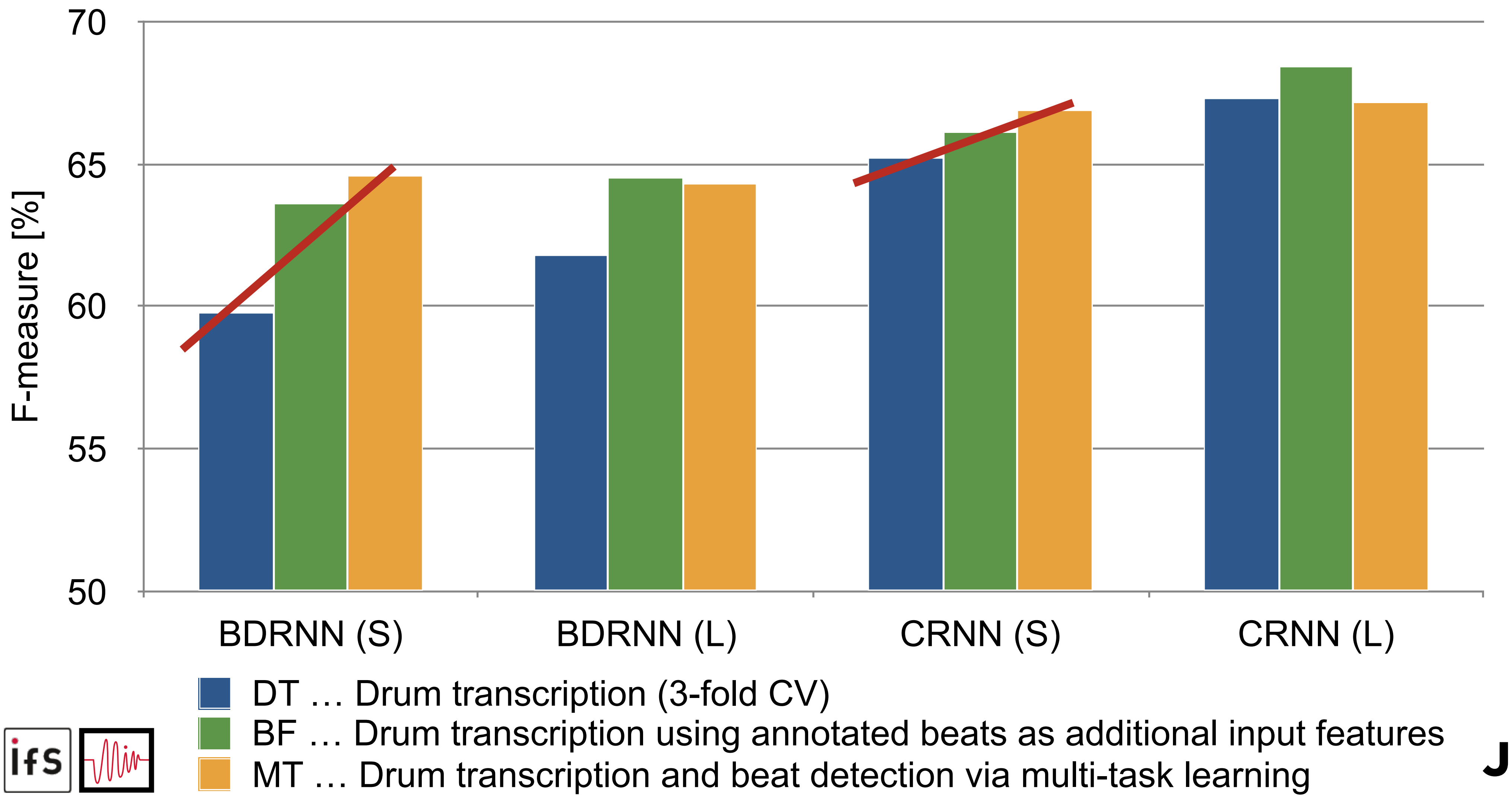


- DT … Drum transcription (3-fold CV)
- BF … Drum transcription using annotated beats as additional input features
- MT … Drum transcription and beat detection via multi-task learning

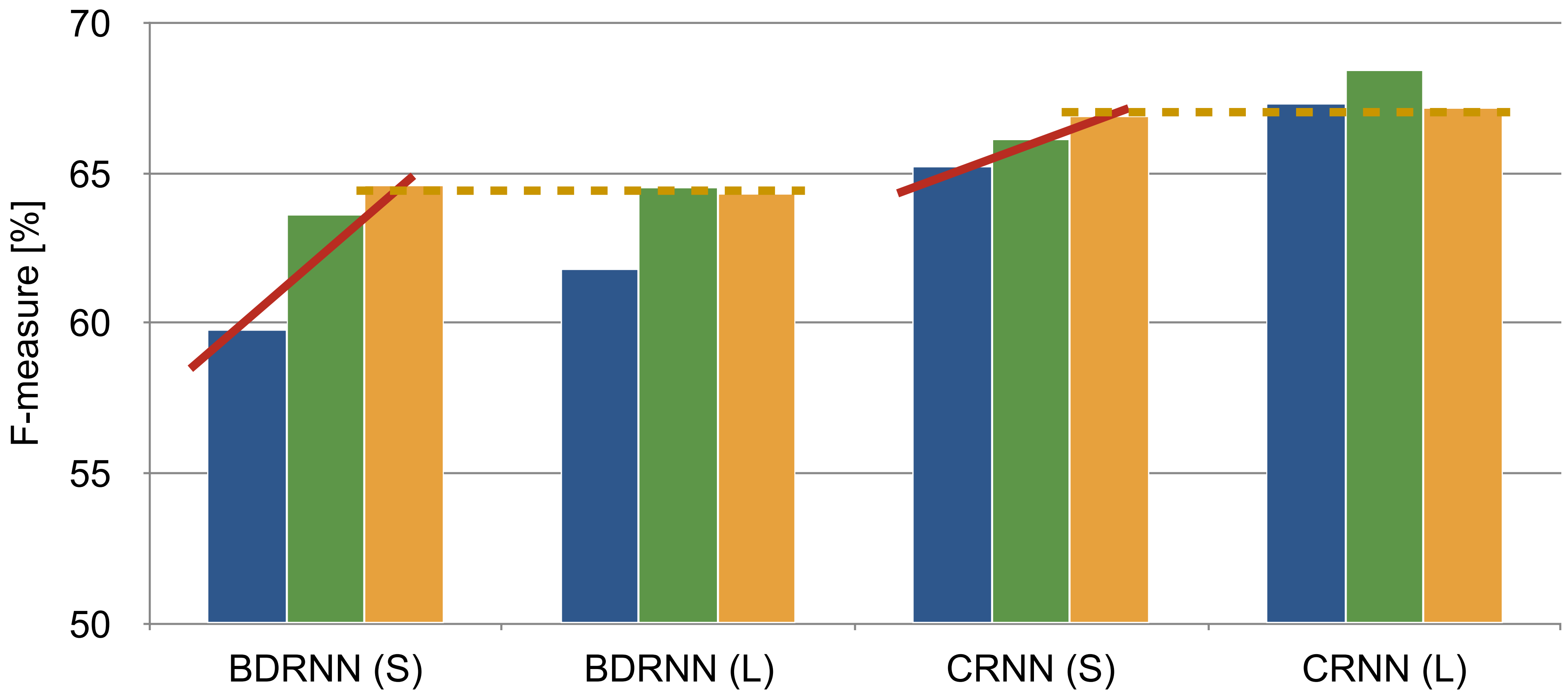
RESULTS FOR RECURRENT ARCHITECTURES



RESULTS FOR RECURRENT ARCHITECTURES

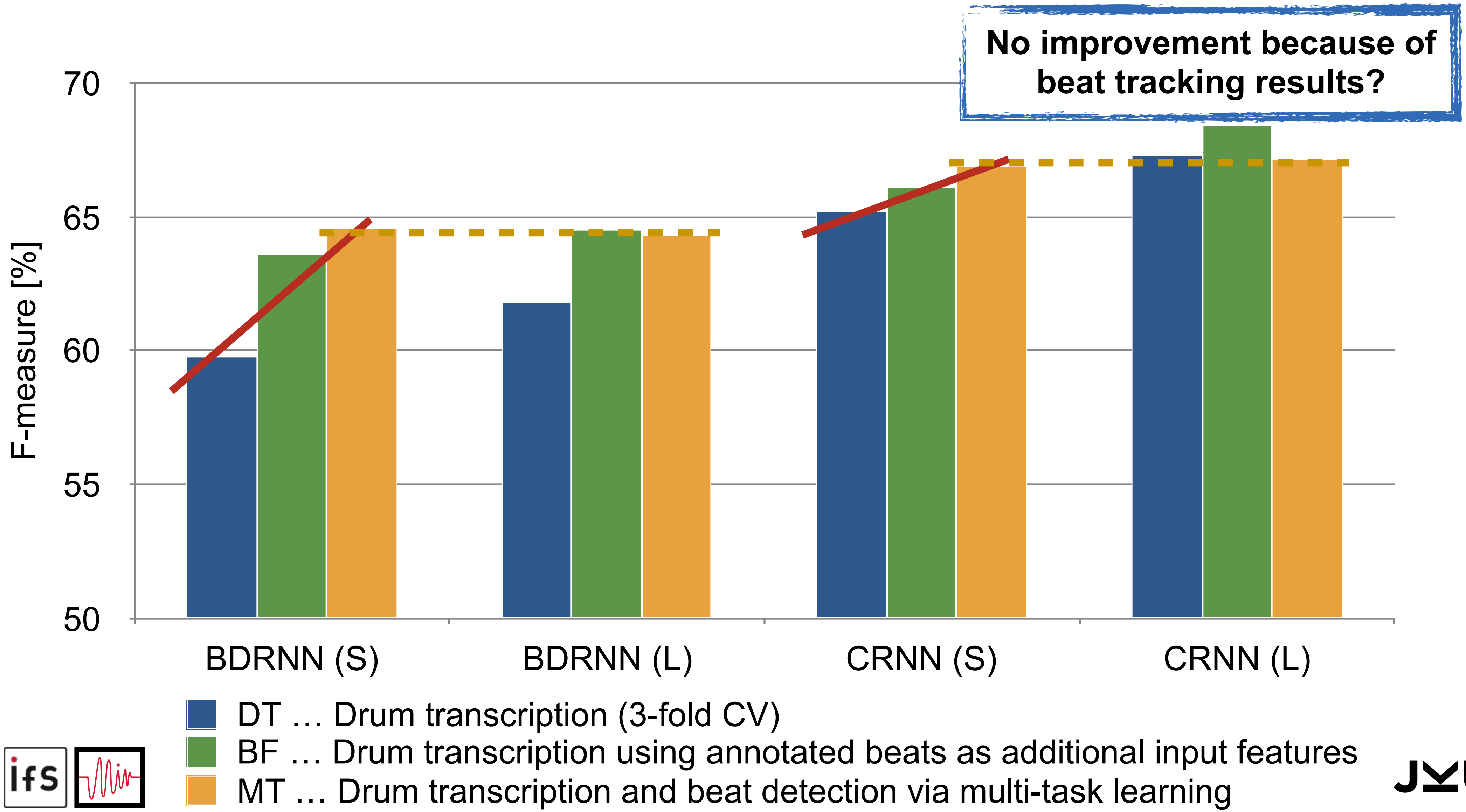


RESULTS FOR RECURRENT ARCHITECTURES



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



CONCLUSIONS

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
  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
  madmom: https://github.com/CPJKU/madmom
- **New dataset** with free music featuring **beat**, and **drum annotations**
  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
  madmom: https://github.com/CPJKU/madmom
- **New dataset** with free music featuring **beat**, and **drum annotations**
  http://ifs.tuwien.ac.at/~vogl/datasets/