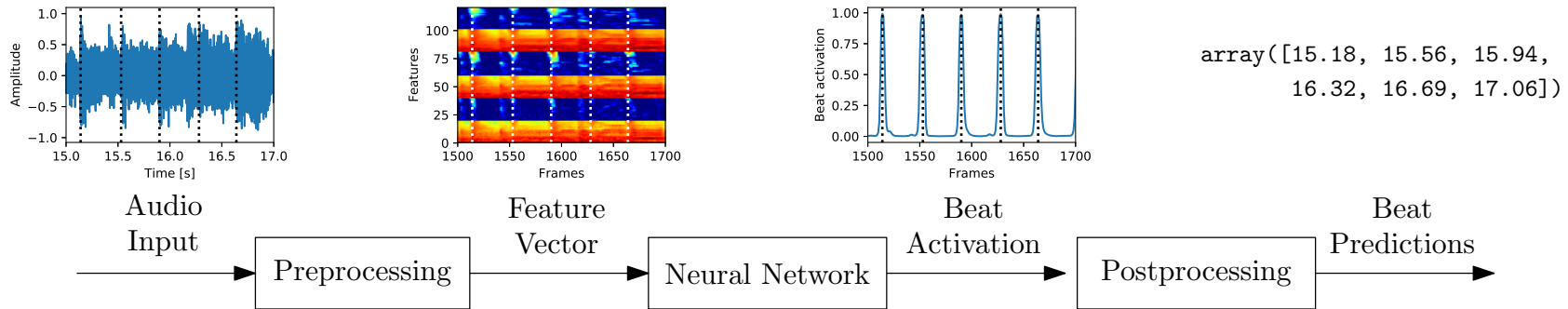


# Multi-Model Beat Tracking based on Recurrent Neural Networks



	Preprocessing	Neural Network	Postprocessing
Methode	Digitale Signalverarbeitung ⇒ DFT, digitale Filter, etc.	Maschinelles Lernen ⇒ Supervised Learning	Stochastisches Modell ⇒ Dynamisches Bayes'sches Netz
Referenz [Böck2014]	<ul style="list-style-type: none"> <li>Log Mel Spektrogram für drei verschiedene Fenstergrößen</li> <li>Median Differenz der Log Mel Spektren</li> </ul>	<ul style="list-style-type: none"> <li>Rekurrentes neuronales Netz (RNN)</li> <li>Bidirektionales LSTM Netzwerk (3 Hidden Layers, Hidden Size 25)</li> <li>Mehrere Netze auf unterschiedliche Datensets trainiert</li> </ul>	<ul style="list-style-type: none"> <li>Hidden Markov Modell mit zwei verborgenden Zuständen: Tempo, Position im Beat</li> <li>Viterbi Algorithmus: Bestimmung der wahrscheinlichsten Sequenz</li> </ul>
Forschungsfragen	<ul style="list-style-type: none"> <li>Verschiedene Feature Repräsentationen ausprobieren: <ul style="list-style-type: none"> <li>Spectral Flux</li> <li>Mel Auditory Feature</li> <li>Harmonic Feature</li> <li>...</li> </ul> </li> <li>Nach welchen Kriterien lassen sich die Datensets in (homogene) Teilmengen unterteilen?</li> </ul>	<ul style="list-style-type: none"> <li>Neue Netzwerk-Architekturen testen z.B.: GRU, CNN, TCN [Bai2018]</li> <li>Hyperparameter Optimierung: <ul style="list-style-type: none"> <li>Learnrate (learning rate schedules)</li> <li>Netzwerkgröße (hidden layers, size)</li> <li>Batch size</li> </ul> </li> <li>Verschiedene Optimierungsmethoden testen: <ul style="list-style-type: none"> <li>Adam, SGD mit Momentum</li> </ul> </li> </ul>	<ul style="list-style-type: none"> <li>Unterschiedliche Zustandsraummodelle testen (Diskretisierung)</li> <li>Parameteroptimierung: <ul style="list-style-type: none"> <li>Übergangswahrscheinlichkeiten</li> <li>Activation Threshold</li> </ul> </li> </ul>

## Datasets:

- Ballroom: 697 Audio Files á 30 Sekunden
  - SMC: 217 Audio Files á 40 Sekunden
  - Hainsworth: 223 Audio Files (25-96 Sekunden)
  - CMR: 118 Audio Files á 120 Sekunden
- } Aufteilung in  $N$  unterschiedliche Teilmengen

## Auswertung:

- 8-fold cross validation beim Trainieren der Netze
- Standard MIR-Metriken (F-measure, P-score, Ccmil, CMLc, CMLt, AMLc, AMLt, Information gain)
- Vergleich mit anderen State-Of-The-Art Beat Trackern (MIREX 2018)
- Submission MIREX 2019 (ISMIR)?