TECHNISCHE UNIVERSITÄT DRESDEN

Europäische Union

Europa fördert Sachsen.
ESF
Europäischer Sozialfonds

Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes.

GHENT UNIVERSITY

Institute of Acoustics and Speech Communication
Chair of Speech Technology and Cognitive Systems

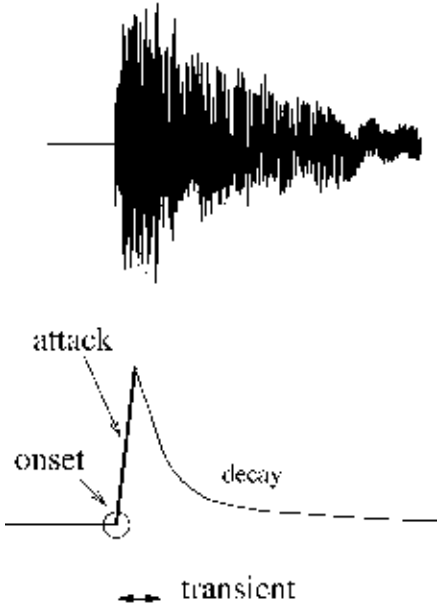
Feature Engineering and Stacked Echo State Networks for Musical Onset Detection

Peter Steiner, Azarakhsh Jalalvand, Simon Stone, Peter Birkholz

ICPR 2020 // 12.01.2021

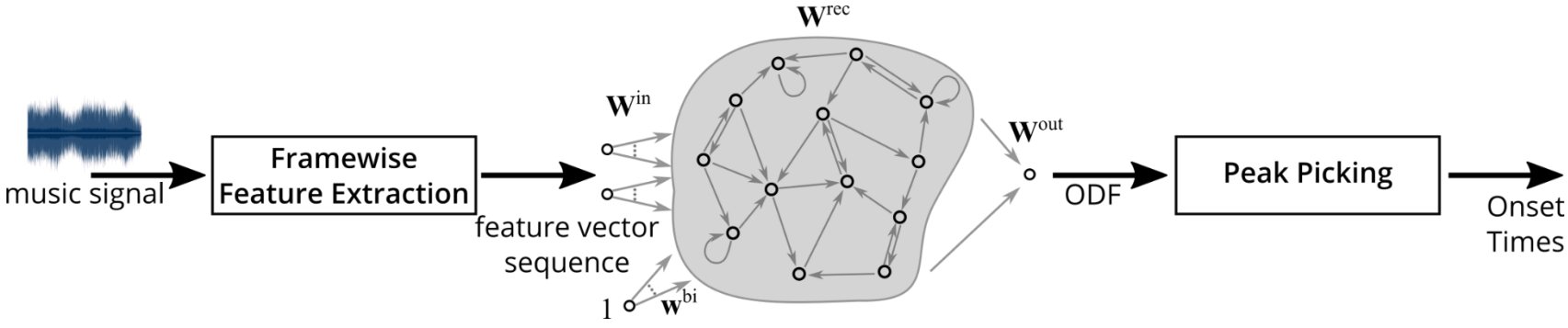
Musical Onset Detection

- Note onset detection:
  - Task of detecting the beginning of new note events in acoustic signals
- Main outline:
  - Transform an audio signal into a Onset Detection Function (ODF)
  - Apply a peak picking algorithm on the ODF to extract onset times



Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005). A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, *13*(5), 1035-1047.

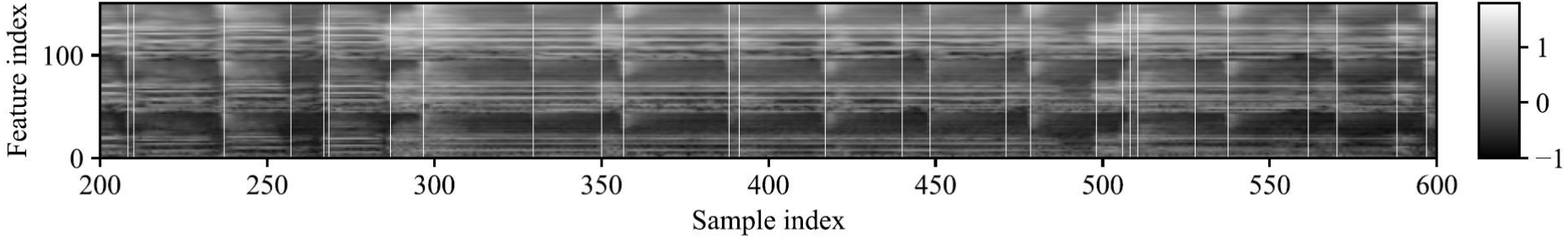
Onset Detection using Echo State Networks
Outline



- Framewise Feature Extraction:
  - Investigate the impact of different window sizes (23 ms, 46 ms, 92 ms)
  - Investigate the impact of different standardization methods
  - Investigate the impact of the second derivative as additional feature
- Echo State Network: Novel way of stacking ESNs to compute the ODF
- Peak Picking: Determine the actual onset times from the ODF

Jaeger, H. (2001). The "echo state" approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, *148* (34), 13.

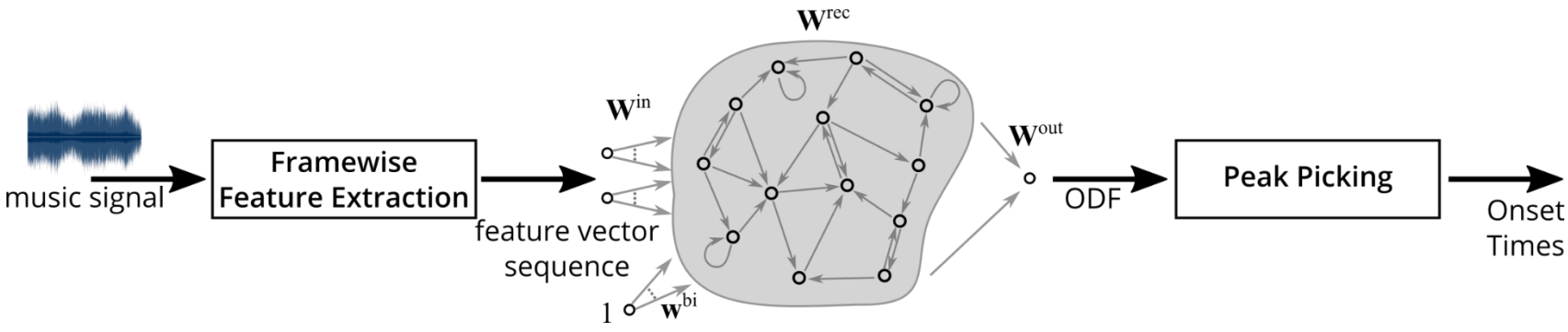
Framewise Feature Extraction



- Top third of the spectrogram was computed with a window of 23 ms
- Medium third of the spectrogram was computed with a window of 46 ms
- Bottom third of the spectrogram was computed with a window of 92 ms

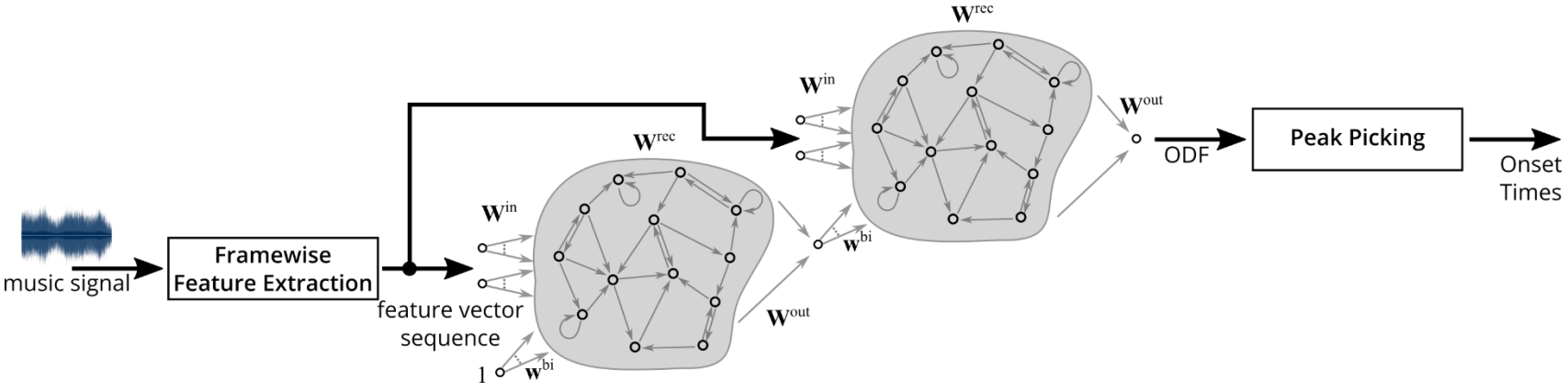
➢ Each part of the spectrogram contains different temporal information to contribute to the onset detection

Onset Detection using Echo State Networks
Outline



- Typically, ESNs can be stacked
  - Output of the first ESN serves as input of the second reservoir
  - Errors from previous layers can be corrected by working with dependencies between outputs of previous layers
  - Here, the ODF is one-dimensional – limited way of error correction

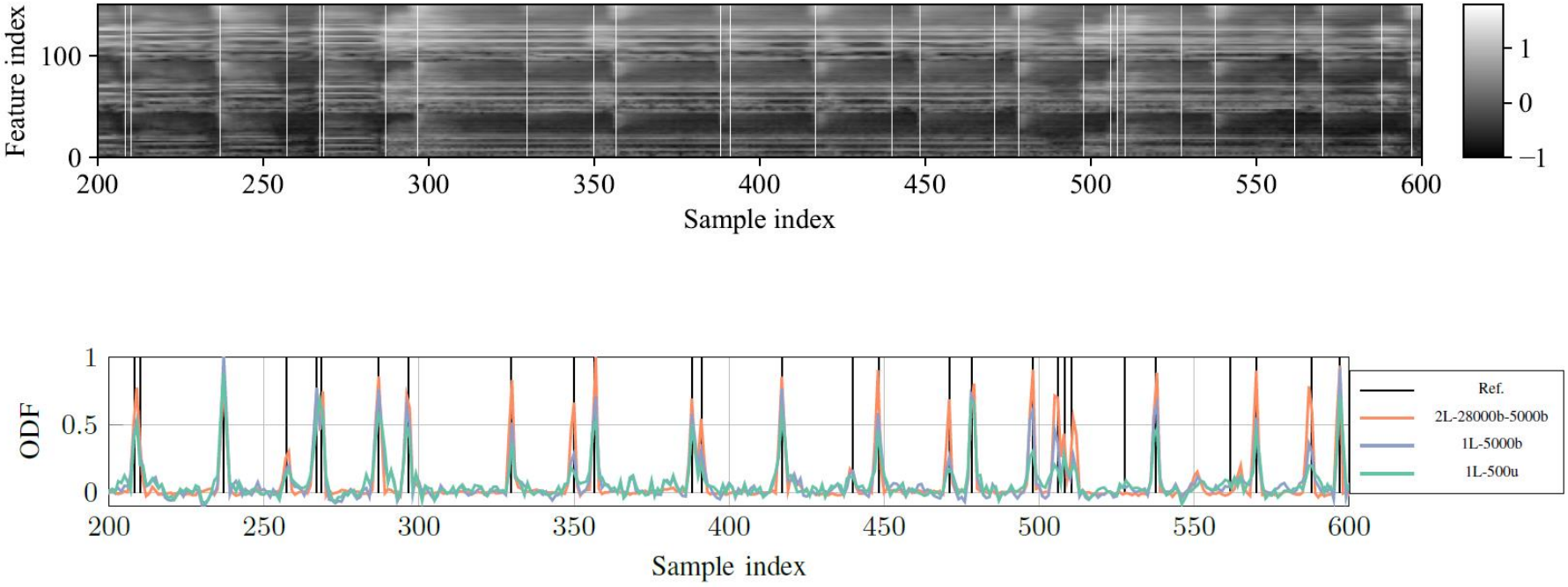
Onset Detection using Echo State Networks
Outline



- Typically, ESNs can be stacked
  - Output of the first ESN serves as input of the second reservoir
  - Errors from previous layers can be corrected by incorporating dependencies between outputs of previous layers
  - Here, the ODF is one-dimensional – limited way of error correction

➢ Use the original feature vector as input and the ODF of the first layer as bias

Results
Example from the test set



- Bidirectional and larger ESNs as well as stacked ESNs all improved the result
- Peaks are getting more prominent and noise is vanishing in the ODF

Results

Architecture	Precision	Recall	F-Measure	Parameters
ESN 1L-24000b	0.881	0.804	0.840	48,001
ESN 2L-28000b-5000b	0.920	0.855	0.886	66,002
ESN (Steiner 2020)	0.854	0.774	0.812	16,001
Bidirectional LSTM (Böck 2010)	0.892	0.855	0.873	20,225
CNN (Schlüter 2013)	0.917	0.889	0.903	289,406

- Promising onset detection results
  - Close to the performance of the state-of-the-art CNN
  - Less trainable parameters than the reference CNNs (66,002 vs. 289,406)
  - Outperformed a bidirectional LSTM

Steiner, P., Stone, S., & Birkholz, P. (2020). Note onset detection using echo state networks. *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2020*, (pp. 157-164).
Eyben, F., Böck, S., Schuller, B., & Graves, A. (2010). Universal onset detection with bidirectional long-short term memory neural networks. In *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010, Utrecht, Netherlands* (pp. 589-594).
Schlüter, J., & Böck, S. (2014, May). Improved musical onset detection with convolutional neural networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6979-6983).

TECHNISCHE UNIVERSITÄT DRESDEN

GHENT UNIVERSITY

Thank you for your attention

QUESTIONS?

Europäische Union

Europa fördert Sachsen.
ESF
Europäischer Sozialfonds

Freistaat Sachsen

Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des vom Sächsischen Landtag beschlossenen Haushaltes.

References

Bello, J. P., Daudet, L., Abdallah, S., Duxbury, C., Davies, M., & Sandler, M. B. (2005). A tutorial on onset detection in music signals. *IEEE Transactions on Speech and Audio Processing*, *13*(5), 1035-1047.

Steiner, P., Stone, S., & Birkholz, P. (2020). Note onset detection using echo state networks. *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2020*, (pp. 157-164).

Jaeger, H. (2001). The “echo state” approach to analysing and training recurrent neural networks-with an erratum note. *Bonn, Germany: German National Research Center for Information Technology GMD Technical Report*, *148* (34), 13.

Eyben, F., Böck, S., Schuller, B., & Graves, A. (2010). Universal onset detection with bidirectional long-short term memory neural networks. In *Proceedings of the 11th International Society for Music Information Retrieval Conference, ISMIR 2010, Utrecht, Netherlands* (pp. 589-594).

Schlüter, J., & Böck, S. (2014, May). Improved musical onset detection with convolutional neural networks. In *2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 6979-6983).