

# Learning to Read and Follow Music in Complete Score Sheet Images

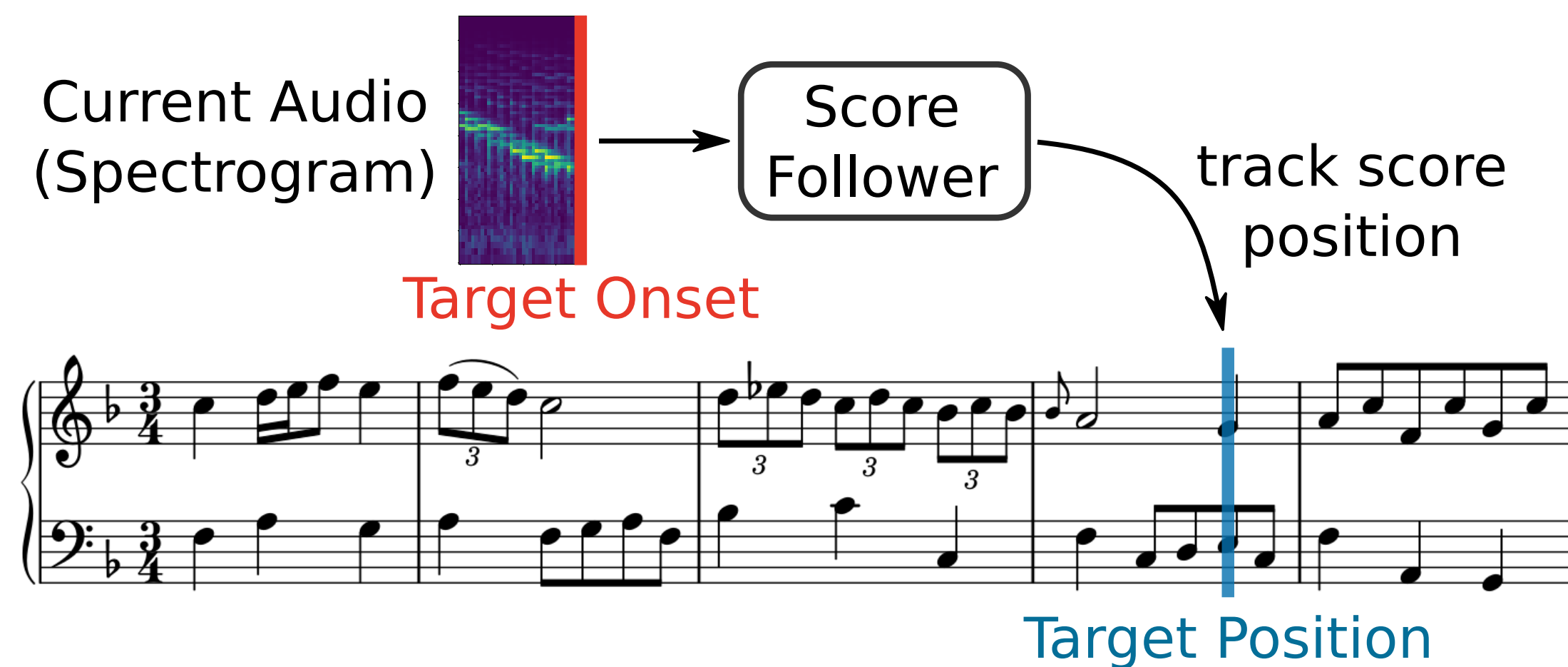
Florian Henkel,<sup>1</sup> Rainer Kelz,<sup>2</sup> Gerhard Widmer<sup>1</sup>

<sup>1</sup>Institute of Computational Perception, JKU

<sup>2</sup>Austrian Research Institute for Artificial Intelligence, OFAI



## Score Following



### DTW / HMM

- + robust and reliable
- need computer-readable score representation

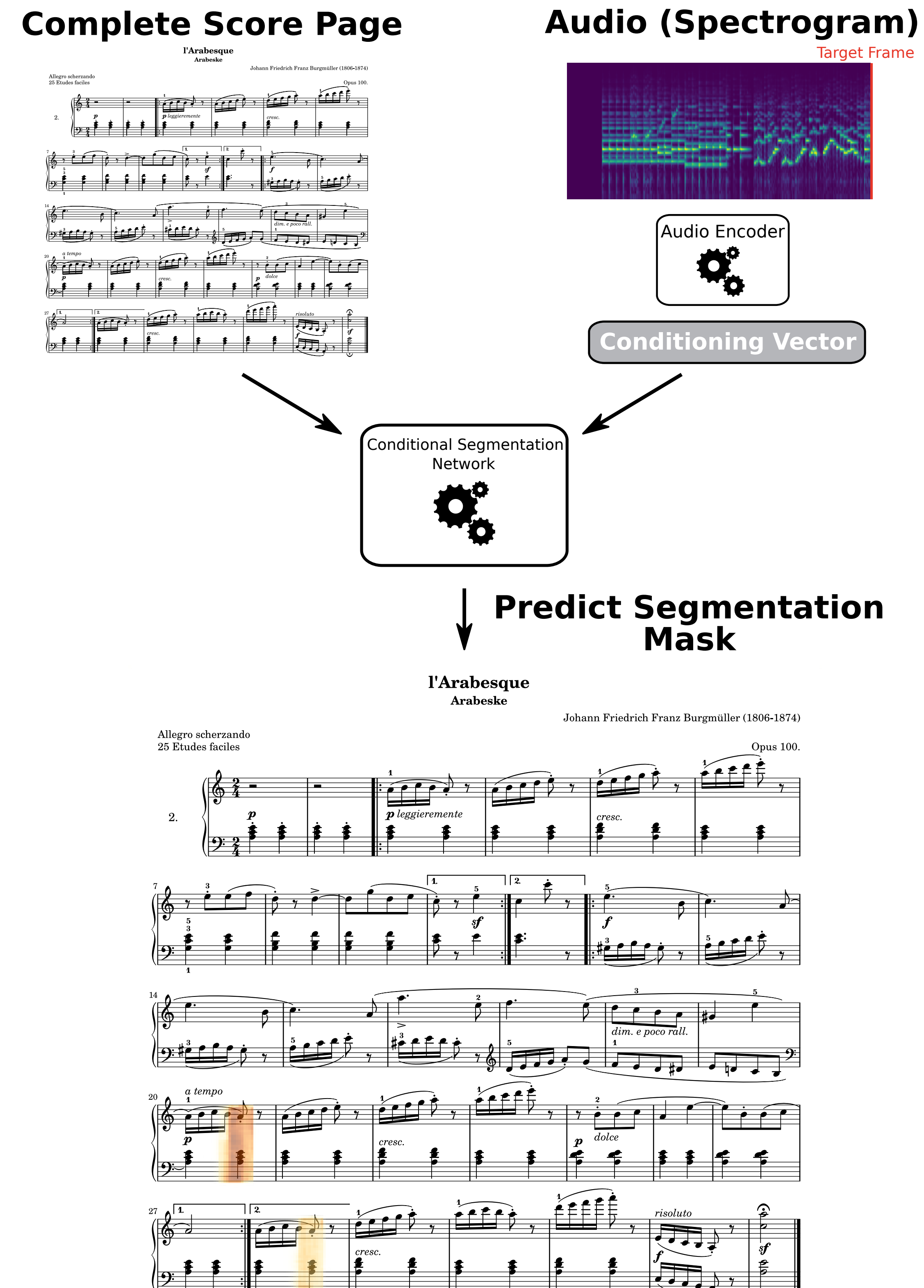
### DNN

- + no computer-readable score representation
- need sheet/audio snippets (score 'unrolling')

### Our proposal

Perform score following directly in complete sheet-image pages (referring image segmentation)

## Approach



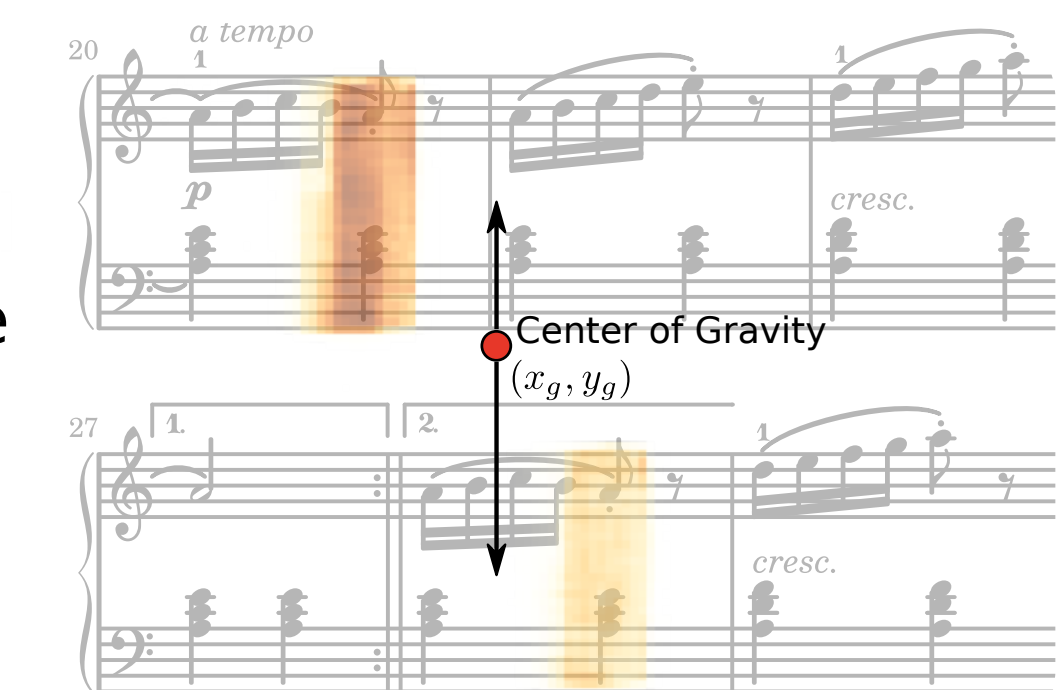
## Evaluation

**Multimodal Sheet Music Dataset (MSMD)[1]**  
synthetic (clean audio/scores), polyphonic piano music

### Evaluation Metric

onsets tracked with error below given threshold

map y-coordinate to closest staff



## Results

MSMD Test-Split

Error [sec]	OMR-ODTW	Dorfer et al. [2]	Henkel et al. [3]	Ours
$\leq 0.05$	44.7%	44.6%	40.9%	<b>73.3%</b>
$\leq 0.10$	51.9%	49.2%	43.3%	<b>74.7%</b>
$\leq 0.50$	76.0%	82.2%	79.7%	<b>85.2%</b>
$\leq 1.00$	85.0%	86.0%	87.8%	<b>88.5%</b>
$\leq 5.00$	<b>97.4%</b>	92.0%	97.2%	93.7%

### Limitations

repeats not handled  
performance degrades for real audio recordings

[1] M. Dorfer, J. Hajič jr., A. Arzt, H. Frostel and G. Widmer. "Learning Audio-Sheet Music Correspondences for Cross-Modal Retrieval and Piece Identification". Transactions of the International Society for Music Information Retrieval, 1(1), 2018

[2] M. Dorfer, A. Arzt, and G. Widmer. "Towards Score Following in Sheet Music Images". In Proc. of the 17th International Society for Music Information Retrieval Conference, New York, USA, 2016

[3] F. Henkel, S. Balke, M. Dorfer, and G. Widmer. "Score Following as a Multi-Modal Reinforcement Learning Problem". Transactions of the International Society for Music Information Retrieval, 2(1), 2019

✉ florian.henkel@jku.at

🐱 [https://github.com/CPJKU/audio\\_conditioned\\_unet](https://github.com/CPJKU/audio_conditioned_unet)

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement number 670035, project "Con Espressione").