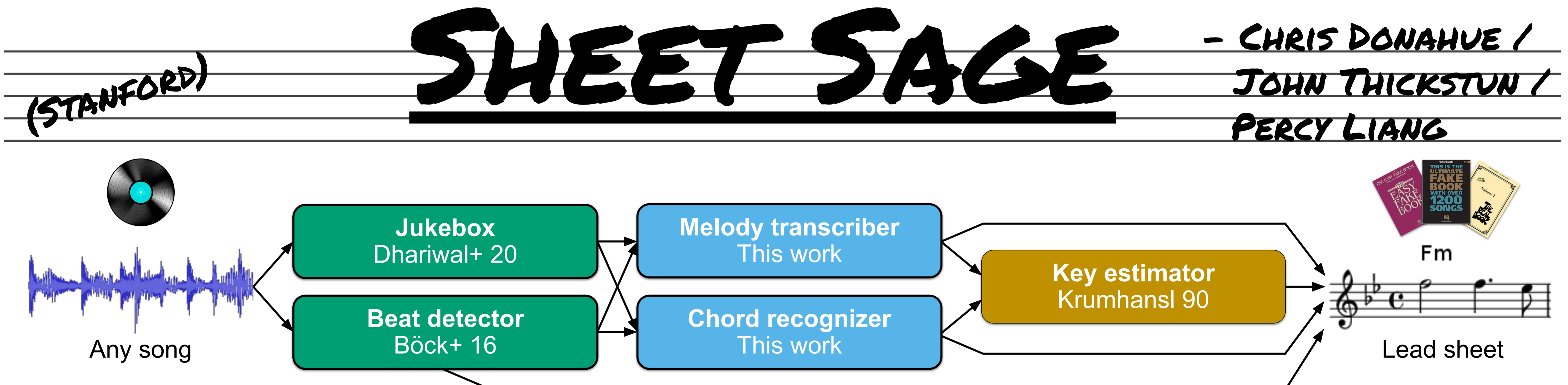
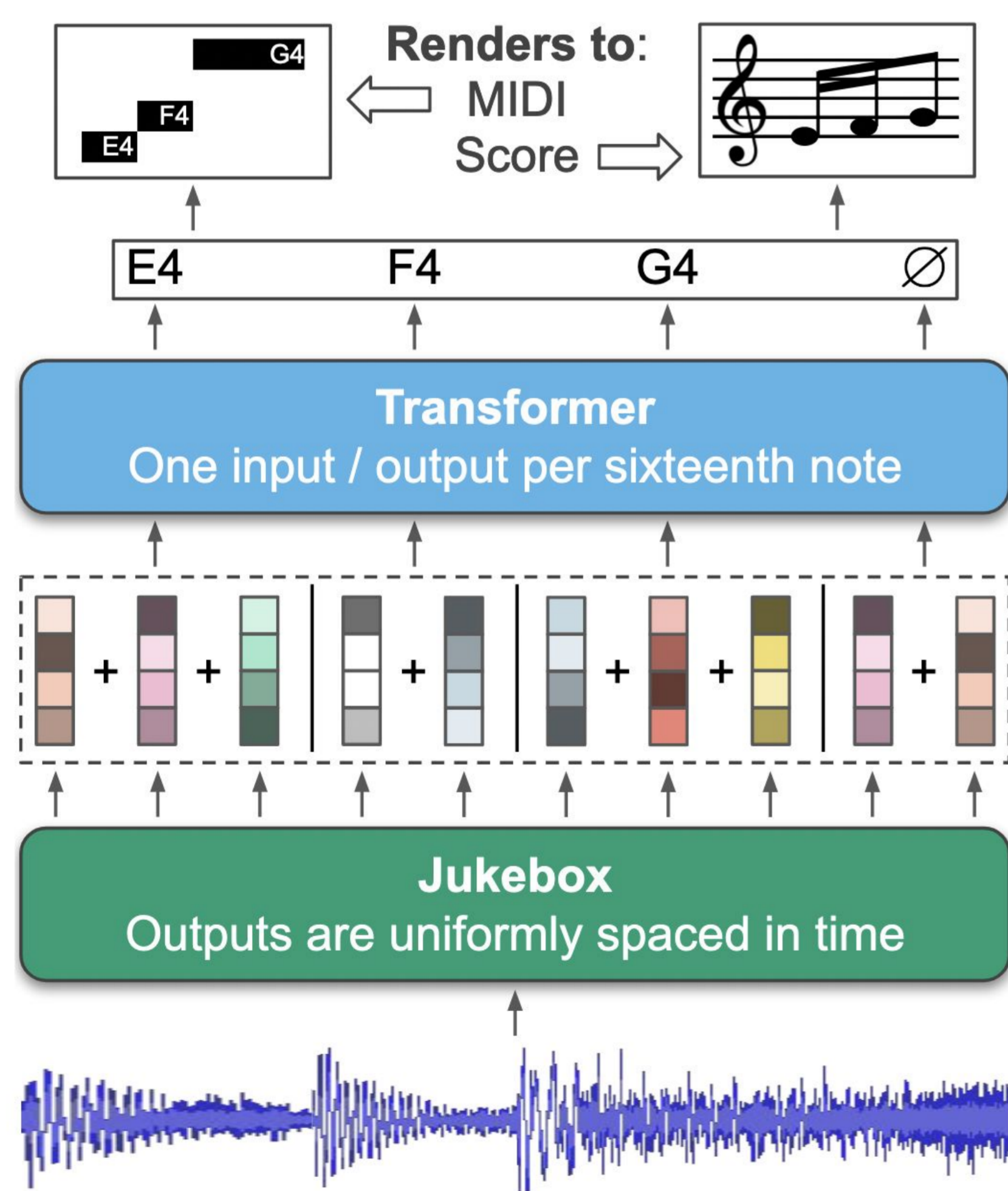


# Improved melody and lead sheet transcription by leveraging generative pre-training on music audio

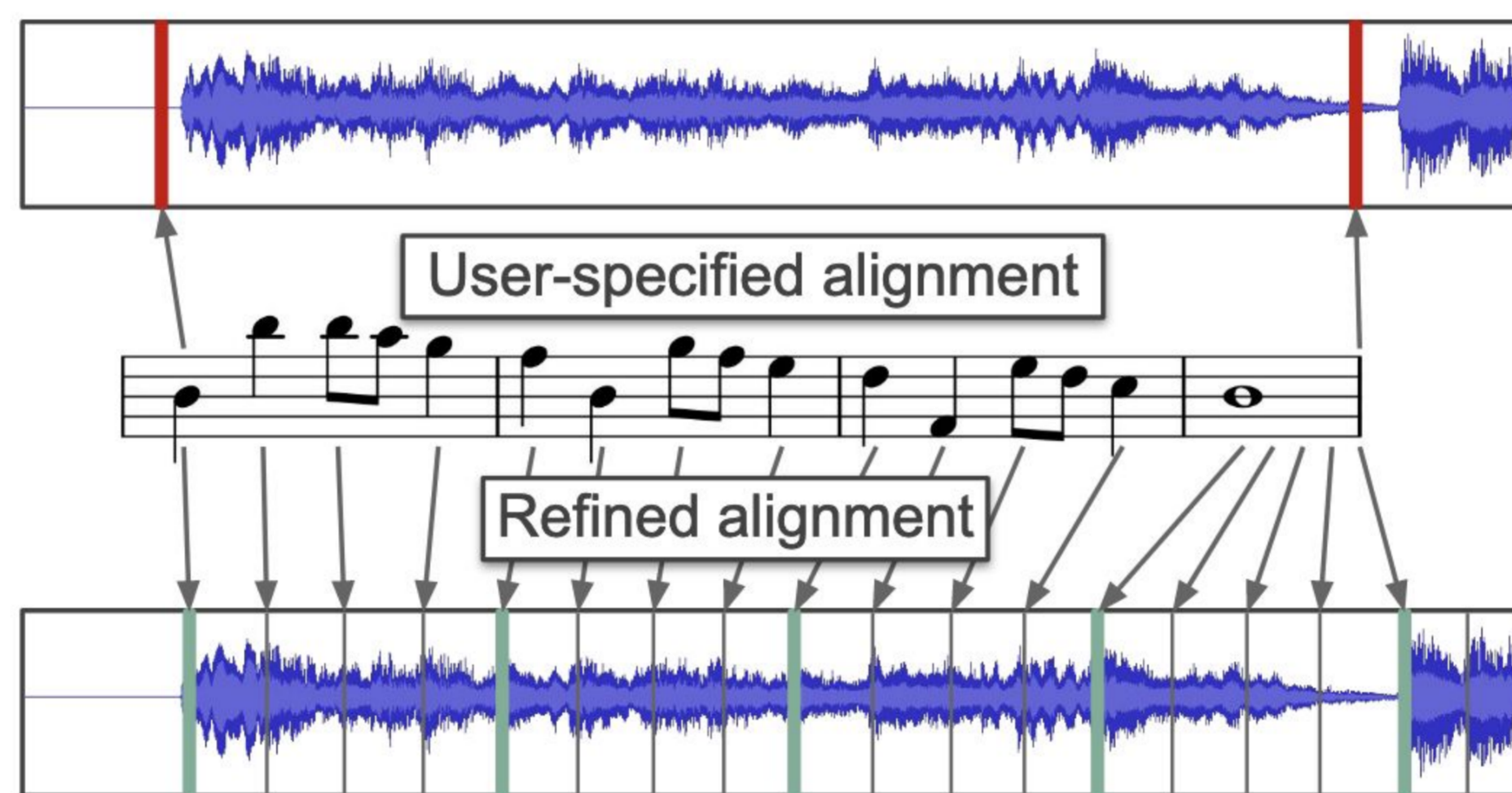


System diagram of **SHEET SAGE**, our approach to transcribing Western music audio into human-readable lead sheet scores.

**SHEET SAGE** is our proposed system for transcribing Western music audio into human-readable lead sheets. Sheet Sage incorporates a new approach to melody transcription:



Our approach uses representations from **Jukebox**, a generative model pre-trained on one million songs. Our previous work (Castellon+ 21) shows that Jukebox is useful for many MIR tasks—this work shows transcription can also benefit.



**SHEET SAGE** is trained on 50+ hours of human-transcribed music lead sheets from [HookTheory](https://hooktheory.com/). We use [madmom](https://github.com/boc/boc) (Böck+ 16) to refine the alignments between the transcriptions and audio. We release these annotations to support future work:

<https://github.com/chrisdonahue/sheetsage>

You can run **SHEET SAGE** on your favorite song using a single command via Docker!

```
> sheetsage.sh input.mp3
```

We compare our melody transcription approach to several baselines using a note-wise, onset-only F1 metric.



We argue that **accurate onset prediction suffices for human-readable melody transcription**.

Method	F <sub>1</sub>
<a href="#">Melody extraction</a> [1] + <a href="#">segment</a>	20
<a href="#">Vocal isolation</a> [2] + <a href="#">transcription</a> [3]	34
<a href="#">DSP features</a> + <a href="#">HMM</a> [4]	42
<a href="#">Spectrogram</a> [5] + Transformer	63
<b>SHEET SAGE</b> (Jukebox + Transformer)	<b>74</b>

Melody transcription performance (note onset F<sub>1</sub>) of our method vs. baselines.

- [1] *Melodia* (Salamon+ 12) w/ ad hoc segmentation
- [2] *Spleeter* (Hennequin+ 20)
- [3] *Tony* (Mauch+ 15)
- [4] HMM-based system from (Ryynänen+ 08)
- [5] Log-mel spectrogram from O&F (Hawthorne+ 17)



[chrisdonahue.com/sheetsage](https://chrisdonahue.com/sheetsage)

**Acknowledgements** Thanks to Annie Hui-Hsin Hsieh, John Hewitt, Maggie Henderson, Megha Srivastava, Nelson Liu, Pang Wei Koh, Rodrigo Castellon, Sam Ainsworth, and Zachary C. Lipton for helpful discussions, support, and advice.

