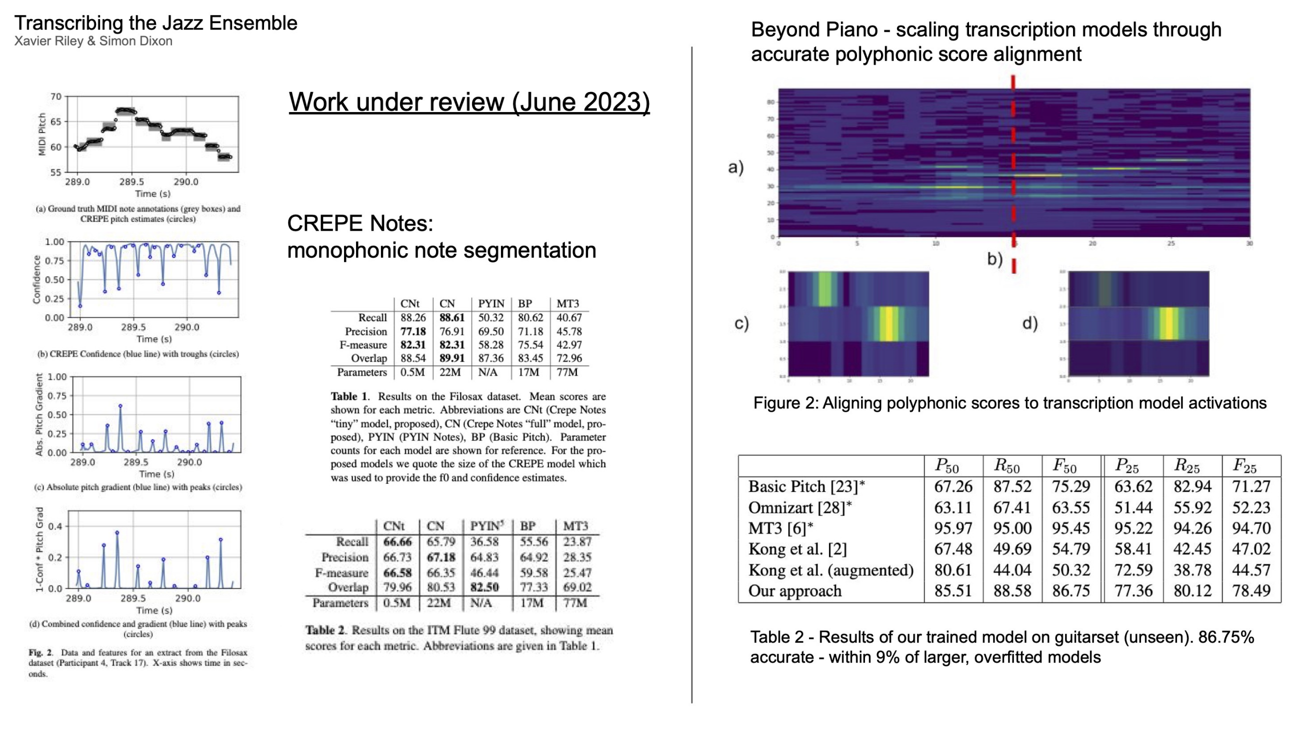
Xavier Riley

Simon Dixon

Transcribing the Jazz Ensemble - towards automatic transcription of small jazz groups

We find that fine grained score alignment accurate enough to train music transcription models. Working with guitar, we trained a model (under review) which achieves SOTA zero shot performance on guitarset with as little as 25ms tolerance.

Is it possible to combine source separation, transcription models and sheet music layout models to transcribe an entire jazz ensemble accurately enough for real consumers?



Transcribing the Jazz Ensemble

Xavier Riley & Simon Dixon

(a) Ground truth MIDI note annotations (grey boxes) and CREPE pitch estimates (circles)

(b) CREPE Confidence (blue line) with troughs (circles)

(c) Absolute pitch gradient (blue line) with peaks (circles)

(d) Combined confidence and gradient (blue line) with peaks (circles)

Fig. 2. Data and features for an extract from the Filosax dataset (Participant 4, Track 17). X-axis shows time in seconds.

Work under review (June 2023)

CREPE Notes:
monophonic note segmentation

	CNt	CN	PYIN	BP	MT3
Recall	88.26	**88.61**	50.32	80.62	40.67
Precision	**77.18**	76.91	69.50	71.18	45.78
F-measure	**82.31**	**82.31**	58.28	75.54	42.97
Overlap	88.54	**89.91**	87.36	83.45	72.96
Parameters	0.5M	22M	N/A	17M	77M

Table 1. Results on the Filosax dataset. Mean scores are shown for each metric. Abbreviations are CNt (Crepe Notes "tiny" model, proposed), CN (Crepe Notes "full" model, proposed), PYIN (PYIN Notes), BP (Basic Pitch). Parameter counts for each model are shown for reference. For the proposed models we quote the size of the CREPE model which was used to provide the f0 and confidence estimates.

	CNt	CN	PYIN[5]	BP	MT3
Recall	**66.66**	65.79	36.58	55.56	23.87
Precision	66.73	**67.18**	64.83	64.92	28.35
F-measure	**66.58**	66.35	46.44	59.58	25.47
Overlap	79.96	80.53	**82.50**	77.33	69.02
Parameters	0.5M	22M	N/A	17M	77M

Table 2. Results on the ITM Flute 99 dataset, showing mean scores for each metric. Abbreviations are given in Table 1.

Beyond Piano - scaling transcription models through accurate polyphonic score alignment

Figure 2: Aligning polyphonic scores to transcription model activations

	P_{50}	R_{50}	F_{50}	P_{25}	R_{25}	F_{25}
Basic Pitch [23]*	67.26	87.52	75.29	63.62	82.94	71.27
Omnizart [28]*	63.11	67.41	63.55	51.44	55.92	52.23
MT3 [6]*	95.97	95.00	95.45	95.22	94.26	94.70
Kong et al. [2]	67.48	49.69	54.79	58.41	42.45	47.02
Kong et al. (augmented)	80.61	44.04	50.32	72.59	38.78	44.57
Our approach	85.51	88.58	86.75	77.36	80.12	78.49

Table 2 - Results of our trained model on guitarset (unseen). 86.75% accurate - within 9% of larger, overfitted models