

## LIST OF TABLES

1	Average raw pitch/chroma accuracies and their standard deviations, tested with the 50 cents threshold.	58
2	Average raw pitch accuracies and their standard deviations, with different evaluation thresholds.	58
3	Train and validation losses as defined in Equation 22 with respect to different variations on the architecture, either limiting or increasing the model capacity. The proposed architecture has the lowest validation loss, indicating that it predicts the most accurate Mel spectrograms while not being prone to overfitting.	80
4	Hyperparameters used during the experiments.	99
5	Frame and note F1 scores are the highest when the non-saturating GAN loss and $\alpha = 0.3$ are used.	100
6	Summary of transcription performance. The non-saturating GAN loss achieves the best performance across all F1 metrics. The average metrics across the tracks in the MAESTRO test dataset are reported, and the model checkpoint where the average of frame F1 and note F1 is the highest on the validation dataset is used.	101
7	A comparison of instrument-agnostic frame transcription accuracies for the baseline models and the proposed synthesizer-aided transcription model. The proposed model achieves better recall while staying at the same precision.	121
8	Per-instrument accuracy of the multi-instrument baseline and the proposed model, showing only the instruments in the test tracks. The multi-instrument baseline generally outperforms the proposed model, which although shows higher precision across all instruments.	123