

Table 1: Inference Result of Onsets and Frames (Train: MAESTRO CNR = ∞ , Test: MAESTRO w/ white noise at various SNRs)

Metric	-6 dB	-3 dB	0 dB	3 dB	6 dB	9 dB	12 dB	15 dB	18 dB	21 dB	24 dB	27 dB	30 dB	33 dB	36 dB	39 dB	42 dB	45 dB	Clean
Precision	92.2 \pm 19.4	94.4 \pm 12.8	95.8 \pm 3.1	95.9 \pm 3.0	96.2 \pm 2.8	96.7 \pm 2.6	97.3 \pm 2.3	97.8 \pm 2.1	98.2 \pm 1.8	98.5 \pm 1.6	98.7 \pm 1.5	98.9 \pm 1.3	99.0 \pm 1.2	99.2 \pm 1.1	99.2 \pm 1.1	99.3 \pm 1.0	99.3 \pm 1.0	99.3 \pm 1.0	99.3 \pm 1.0
Recall	18.5 \pm 13.5	33.1 \pm 18.0	48.4 \pm 20.4	62.0 \pm 19.8	72.5 \pm 17.0	79.2 \pm 14.4	83.4 \pm 12.1	86.0 \pm 10.8	87.5 \pm 10.1	88.4 \pm 9.7	89.0 \pm 9.4	89.3 \pm 9.2	89.5 \pm 9.1	89.8 \pm 8.8	89.9 \pm 8.9	90.1 \pm 8.7	90.1 \pm 8.7	90.2 \pm 8.7	90.5 \pm 8.2
F1 Score	28.9 \pm 18.1	46.4 \pm 20.5	61.6 \pm 19.9	73.4 \pm 16.9	81.5 \pm 13.2	86.3 \pm 10.6	89.4 \pm 8.6	91.1 \pm 7.5	92.2 \pm 6.9	92.9 \pm 6.5	93.4 \pm 6.4	93.6 \pm 6.1	93.8 \pm 6.0	94.0 \pm 5.7	94.1 \pm 5.8	94.2 \pm 5.6	94.3 \pm 5.6	94.3 \pm 5.6	94.5 \pm 5.2

Table 2: Inference Result of Kong et al. (Train: MAESTRO CNR = ∞ , Test: MAESTRO w/ white noise at various SNRs)

Metric	-6 dB	-3 dB	0 dB	3 dB	6 dB	9 dB	12 dB	15 dB	18 dB	21 dB	24 dB	27 dB	30 dB	33 dB	36 dB	39 dB	42 dB	45 dB	Clean
Precision	68.0 \pm 41.5	79.4 \pm 31.4	86.8 \pm 20.0	90.0 \pm 9.2	91.3 \pm 5.4	92.7 \pm 4.8	94.1 \pm 4.4	95.4 \pm 4.0	96.3 \pm 3.7	96.8 \pm 3.5	97.4 \pm 3.0	97.7 \pm 2.7	97.8 \pm 2.6	98.0 \pm 2.4	98.1 \pm 2.3	98.2 \pm 2.2	98.2 \pm 2.2	98.2 \pm 2.2	98.0 \pm 2.3
Recall	9.2 \pm 12.2	20.4 \pm 18.3	36.9 \pm 23.6	55.4 \pm 25.4	71.2 \pm 22.6	82.3 \pm 17.1	88.9 \pm 12.0	92.2 \pm 8.6	93.8 \pm 7.0	94.6 \pm 6.1	95.1 \pm 5.6	95.4 \pm 5.0	95.7 \pm 4.7	95.9 \pm 4.5	96.0 \pm 4.4	96.1 \pm 4.4	96.1 \pm 4.3	96.2 \pm 4.2	96.4 \pm 4.0
F1 Score	14.6 \pm 17.1	29.6 \pm 22.9	47.7 \pm 25.2	64.5 \pm 23.7	77.4 \pm 18.5	86.1 \pm 12.8	91.0 \pm 8.8	93.6 \pm 6.3	94.9 \pm 5.2	95.6 \pm 4.7	96.2 \pm 4.2	96.5 \pm 3.7	96.7 \pm 3.5	96.9 \pm 3.3	97.0 \pm 3.2	97.1 \pm 3.1	97.1 \pm 3.1	97.2 \pm 3.1	97.2 \pm 3.0

Table 3: Inference Result of Onsets and Frames (Train: MAESTRO CNR = 1, Test: MAESTRO w/ white noise at various SNRs)

Metric	-6 dB	-3 dB	0 dB	3 dB	6 dB	9 dB	12 dB	15 dB	18 dB	21 dB	24 dB	27 dB	30 dB	33 dB	36 dB	39 dB	42 dB	45 dB	Clean
Precision	94.9 \pm 3.4	96.0 \pm 2.9	97.1 \pm 2.4	97.7 \pm 2.2	98.2 \pm 2.0	98.4 \pm 1.8	98.6 \pm 1.7	98.7 \pm 1.6	98.8 \pm 1.5	98.9 \pm 1.4	99.0 \pm 1.3	99.0 \pm 1.2	99.1 \pm 1.2	99.1 \pm 1.2	99.1 \pm 1.1	99.1 \pm 1.2	99.1 \pm 1.2	99.1 \pm 1.2	99.1 \pm 1.2
Recall	62.0 \pm 18.6	71.8 \pm 16.8	78.7 \pm 14.4	83.0 \pm 12.6	85.7 \pm 11.2	87.3 \pm 10.3	88.3 \pm 9.6	88.9 \pm 9.1	89.4 \pm 8.7	89.8 \pm 8.5	90.0 \pm 8.3	90.2 \pm 8.2	90.4 \pm 8.2	90.5 \pm 8.2	90.5 \pm 8.1	90.6 \pm 8.0	90.6 \pm 8.0	90.7 \pm 7.9	90.7 \pm 7.9
F1 Score	73.5 \pm 15.6	81.1 \pm 12.9	86.2 \pm 10.4	89.3 \pm 8.8	91.1 \pm 7.6	92.2 \pm 7.0	92.9 \pm 6.4	93.3 \pm 5.9	93.7 \pm 5.6	93.9 \pm 5.5	94.1 \pm 5.3	94.2 \pm 5.2	94.4 \pm 5.2	94.4 \pm 5.2	94.5 \pm 5.1	94.5 \pm 5.1	94.5 \pm 5.0	94.5 \pm 5.0	94.6 \pm 5.0

Table 4: Inference Result of Onsets and Frames (Train: MAESTRO CNR = 3, Test: MAESTRO w/ white noise at various SNRs)

Metric	-6 dB	-3 dB	0 dB	3 dB	6 dB	9 dB	12 dB	15 dB	18 dB	21 dB	24 dB	27 dB	30 dB	33 dB	36 dB	39 dB	42 dB	45 dB	Clean
Precision	94.8 \pm 4.0	96.2 \pm 2.7	97.3 \pm 2.1	97.8 \pm 2.0	98.2 \pm 1.8	98.5 \pm 1.6	98.7 \pm 1.5	98.8 \pm 1.3	98.9 \pm 1.3	99.0 \pm 1.2	99.1 \pm 1.1	99.1 \pm 1.1	99.2 \pm 1.1	99.2 \pm 1.1	99.2 \pm 1.1	99.2 \pm 1.0	99.3 \pm 1.0	99.2 \pm 1.0	99.2 \pm 1.0
Recall	56.4 \pm 20.7	68.0 \pm 18.6	76.2 \pm 15.5	81.3 \pm 13.3	84.3 \pm 11.7	86.5 \pm 10.7	87.7 \pm 10.0	88.6 \pm 9.5	89.3 \pm 9.0	89.7 \pm 8.8	89.9 \pm 8.7	90.2 \pm 8.4	90.4 \pm 8.3	90.4 \pm 8.2	90.5 \pm 8.2	90.6 \pm 8.1	90.7 \pm 8.1	90.8 \pm 8.1	91.1 \pm 7.9
F1 Score	68.5 \pm 18.6	78.2 \pm 14.8	84.6 \pm 11.3	88.2 \pm 9.4	90.3 \pm 8.1	91.8 \pm 7.2	92.6 \pm 6.6	93.2 \pm 6.2	93.6 \pm 5.8	93.9 \pm 5.7	94.1 \pm 5.5	94.3 \pm 5.3	94.4 \pm 5.3	94.4 \pm 5.2	94.5 \pm 5.2	94.5 \pm 5.1	94.6 \pm 5.1	94.6 \pm 5.1	94.8 \pm 4.9

Table 5: Inference Result of Onsets and Frames (Train: MAESTRO CNR = 1/3, Test: MAESTRO w/ white noise at various SNRs)

Metric	-6 dB	-3 dB	0 dB	3 dB	6 dB	9 dB	12 dB	15 dB	18 dB	21 dB	24 dB	27 dB	30 dB	33 dB	36 dB	39 dB	42 dB	45 dB	Clean
Precision	95.1 \pm 3.3	96.4 \pm 2.7	97.3 \pm 2.1	98.0 \pm 1.9	98.4 \pm 1.7	98.6 \pm 1.6	98.7 \pm 1.6	98.8 \pm 1.5	98.9 \pm 1.4	99.0 \pm 1.4	99.0 \pm 1.3	99.1 \pm 1.2	99.1 \pm 1.2	99.1 \pm 1.2	99.2 \pm 1.1	99.1 \pm 1.1	99.1 \pm 1.2	99.1 \pm 1.1	99.1 \pm 1.2
Recall	62.4 \pm 19.4	72.3 \pm 16.9	79.0 \pm 14.2	83.0 \pm 12.4	85.5 \pm 11.3	87.0 \pm 10.5	87.9 \pm 9.9	88.4 \pm 9.4	88.9 \pm 9.2	89.2 \pm 9.0	89.5 \pm 8.8	89.5 \pm 8.7	89.7 \pm 8.6	89.7 \pm 8.6	89.8 \pm 8.6	89.8 \pm 8.5	89.9 \pm 8.5	89.9 \pm 8.5	90.0 \pm 8.4
F1 Score	73.6 \pm 16.3	81.5 \pm 13.0	86.5 \pm 10.3	89.4 \pm 8.8	91.1 \pm 7.8	92.1 \pm 7.1	92.7 \pm 6.6	93.1 \pm 6.2	93.4 \pm 6.1	93.6 \pm 5.9	93.8 \pm 5.7	93.8 \pm 5.6	93.9 \pm 5.6	94.0 \pm 5.6	94.1 \pm 5.5	94.1 \pm 5.5	94.1 \pm 5.5	94.1 \pm 5.5	94.2 \pm 5.4

Table 6: Inference Result of Onsets and Frames (Train: MAESTRO CNR = 0, Test: MAESTRO w/ white noise at various SNRs)

Metric	-6 dB	-3 dB	0 dB	3 dB	6 dB	9 dB	12 dB	15 dB	18 dB	21 dB	24 dB	27 dB	30 dB	33 dB	36 dB	39 dB	42 dB	45 dB	Clean
Precision	97.3 \pm 2.4	97.9 \pm 2.1	98.5 \pm 1.5	98.8 \pm 1.3	99.0 \pm 1.2	99.2 \pm 0.9	99.2 \pm 1.0	99.3 \pm 1.0	99.3 \pm 0.9	99.3 \pm 1.0	99.3 \pm 0.9	99.3 \pm 0.9	99.3 \pm 1.0	99.3 \pm 0.9	99.3 \pm 1.0	99.3 \pm 1.0	99.3 \pm 1.0	99.2 \pm 1.0	99.2 \pm 1.0
Recall	54.9 \pm 18.9	66.1 \pm 17.6	73.8 \pm 15.7	79.1 \pm 14.2	82.5 \pm 12.8	84.4 \pm 11.8	85.5 \pm 11.2	86.3 \pm 10.8	86.8 \pm 10.4	87.2 \pm 10.2	87.5 \pm 10.0	87.7 \pm 9.9	87.9 \pm 9.8	88.1 \pm 9.6	88.2 \pm 9.6	88.3 \pm 9.4	88.4 \pm 9.4	88.4 \pm 9.3	88.6 \pm 9.3
F1 Score	68.3 \pm 17.0	77.5 \pm 14.3	83.4 \pm 11.9	87.1 \pm 10.2	89.4 \pm 8.9	90.7 \pm 7.9	91.5 \pm 7.5	92.0 \pm 7.1	92.3 \pm 6.8	92.5 \pm 6.7	92.7 \pm 6.5	92.8 \pm 6.4	92.9 \pm 6.3	93.1 \pm 6.1	93.1 \pm 6.1	93.2 \pm 6.1	93.2 \pm 6.0	93.3 \pm 6.0	93.3 \pm 5.9