

# APPENDIX

## FULL PAIRWISE ANALYSIS TABLES

TABLE I

PAIRWISE DIFFERENCES OF SELECTED SYSTEMS. CREDIBILITY (CREDIB.) IS PROBABILITY OF IMPROVEMENT (OF THE BETTER SYSTEM) MEASURED BY BOOTSTRAP ESTIMATE [1]. SENTENCE DIFFERENCE (S-DIFF.) IS THE PERCENTAGE OF TRANSCRIPTIONS THAT WERE DIFFERENT BETWEEN A AND B.

	A			B			WER Difference (B-A)			S-Diff.	WER Rank $\tau$	
	Family	Encoder	LM Data	Family	Encoder	LM Data	[pp.]	Relative	Credib.	[%]	Utt.	Spk.
FP Train20 $\rightarrow$ FP Test16	HMM	CRDNN	Transcript	AED	CRDNN	-	+2.18	+21%	100%	63.85	0.54	0.87
	HMM	CRDNN	Parl30M	AED	CRDNN	Parl30M	+0.72	+11%	100%	48.18	0.59	0.73
	FS HMM	CRDNN	Transcript	AED	CRDNN	-	-0.16	-1.6%	79%	72.67	0.47	0.80
	FS HMM	CRDNN	Parl30M	AED	CRDNN	Parl30M	+0.13	+2.0%	79%	53.48	0.54	0.83
	HMM	w2v2.0	Transcript	AED/CTC	w2v2.0	-	+0.32	+4.8%	99%	51.09	0.59	0.84
	HMM	w2v2.0	Parl30M	AED/CTC	w2v2.0	Parl30M	+0.09	+1.7%	79%	42.84	0.65	0.87
	AED	CRDNN	Parl30M	AED	CRDNN	-	+3.72	+36%	100%	53.32	0.62	0.70
	AED/CTC	w2v2.0	-	AED	CRDNN	-	+3.60	+35%	100%	64.83	0.51	0.73
	AED/CTC	w2v2.0	Parl30M	AED	CRDNN	Parl30M	+0.63	+9.6%	100%	45.49	0.60	0.78
	AED/CTC	w2v2.0	Parl30M	AED/CTC	w2v2.0	-	+0.76	+11%	100%	37.81	0.71	0.82
	HMM	CRDNN	Parl30M	HMM	CRDNN	Transcript	+2.27	+28%	100%	50.67	0.63	0.68
	HMM	w2v2.0	Transcript	HMM	CRDNN	Transcript	+1.75	+21%	100%	52.96	0.59	0.78
	HMM	w2v2.0	Parl30M	HMM	CRDNN	Parl30M	+0.01	+0.2%	54%	40.04	0.66	0.88
	HMM	w2v2.0	Parl30M	HMM	w2v2.0	Transcript	+0.53	+8.3%	100%	40.04	0.71	0.85
	HMM	CRDNN	Transcript	FS HMM	CRDNN	Transcript	+2.34	+22%	100%	60.74	0.57	0.86
	HMM	CRDNN	Parl30M	FS HMM	CRDNN	Parl30M	+0.59	+9.0%	100%	42.32	0.65	0.83
	FS HMM	CRDNN	Parl30M	FS HMM	CRDNN	Transcript	+4.02	+38%	100%	63.07	0.51	0.78
FP Train20 $\rightarrow$ FP Test20	HMM	CRDNN	Transcript	AED	CRDNN	-	+0.98	+11%	100%	38.11	0.60	0.84
	HMM	CRDNN	Parl30M	AED	CRDNN	Parl30M	-0.83	-11%	100%	36.52	0.63	0.79
	FS HMM	CRDNN	Transcript	AED	CRDNN	-	-0.82	-9.5%	100%	46.28	0.53	0.81
	FS HMM	CRDNN	Parl30M	AED	CRDNN	Parl30M	-2.23	-30%	100%	42.37	0.58	0.77
	HMM	w2v2.0	Transcript	AED/CTC	w2v2.0	-	+0.12	+1.8%	82%	28.31	0.68	0.85
	HMM	w2v2.0	Parl30M	AED/CTC	w2v2.0	Parl30M	-0.93	-13%	100%	34.28	0.66	0.80
	AED	CRDNN	Parl30M	AED	CRDNN	-	+1.17	+13%	100%	30.95	0.70	0.78
	AED/CTC	w2v2.0	-	AED	CRDNN	-	+1.97	+23%	100%	38.69	0.57	0.77
	AED/CTC	w2v2.0	Parl30M	AED	CRDNN	Parl30M	+0.30	+4.1%	96%	32.22	0.66	0.84
	AED/CTC	w2v2.0	Parl30M	AED/CTC	w2v2.0	-	-0.49	-7.4%	100%	22.85	0.77	0.87
	HMM	CRDNN	Parl30M	HMM	CRDNN	Transcript	-0.64	-8.5%	100%	36.91	0.64	0.86
	HMM	w2v2.0	Transcript	HMM	CRDNN	Transcript	+1.11	+15%	100%	30.60	0.66	0.81
	HMM	w2v2.0	Parl30M	HMM	CRDNN	Parl30M	+0.21	+2.5%	92%	32.15	0.70	0.79
	HMM	w2v2.0	Parl30M	HMM	w2v2.0	Transcript	-1.54	-24%	100%	31.49	0.68	0.83
	HMM	CRDNN	Transcript	FS HMM	CRDNN	Transcript	+1.80	+19%	100%	36.83	0.64	0.80
	HMM	CRDNN	Parl30M	FS HMM	CRDNN	Parl30M	+1.40	+15%	100%	39.00	0.63	0.82
	FS HMM	CRDNN	Parl30M	FS HMM	CRDNN	Transcript	-0.24	-2.6%	91%	45.47	0.56	0.72

## REFERENCES

- [1] M. Bisani and H. Ney, "Bootstrap estimates for confidence intervals in asr performance evaluation," in *2004 IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 1, 2004, pp. 1-409.

TABLE II

PAIRWISE DIFFERENCES OF SELECTED SYSTEMS. CREDIBILITY (CREDIB.) IS PROBABILITY OF IMPROVEMENT (OF THE BETTER SYSTEM) MEASURED BY BOOTSTRAP ESTIMATE [1]. SENTENCE DIFFERENCE (S-DIFF.) IS THE PERCENTAGE OF TRANSCRIPTIONS THAT WERE DIFFERENT BETWEEN A AND B. THE SUPERSCRIPTS  $^{MW}$  AND  $^{FT}$  MEAN THE POST-10-MODEL-AVERAGING MINIMUM WORD ERROR RATE FINETUNING AND HMM CONTINUED TRAINING, RESPECTIVELY.

	A			B			WER Difference (B-A)			S-Diff.	WER Rank $\tau$	
	Family	Encoder	LM Data	Family	Encoder	LM Data	[pp.]	Relative	Credib.	[%]	Utt.	Spk.
Librispeech Test Clean	HMM	CRDNN	Transcript	AED	CRDNN	-	+0.55	+11%	100%	47.60	0.51	0.69
	HMM	CRDNN	Orig. 800M	AED	CRDNN	-	+1.22	+24%	100%	47.71	0.45	0.53
	FS HMM	CRDNN	Transcript	AED	CRDNN	-	-0.24	-4.7%	97%	51.15	0.49	0.65
	FS HMM	CRDNN	Orig. 800M	AED	CRDNN	-	+0.70	+14%	100%	51.18	0.42	0.47
	HMM	w2v2.0	Transcript	AED/CTC	w2v2.0	-	+0.17	+7.0%	100%	26.95	0.60	0.60
	HMM	Conformer <sub>S</sub>	Transcript	AED/CTC	Conformer <sub>S</sub>	-	+0.05	+1.4%	72%	35.95	0.58	0.64
	HMM	Conformer <sub>S</sub>	Orig. 800M	AED/CTC	Conformer <sub>S</sub>	-	+0.28	+7.7%	98%	38.85	0.50	0.52
	HMM	Conformer <sub>S</sub> <sup>FT</sup>	Transcript	AED/CTC	Conformer <sub>S</sub> <sup>MW</sup>	-	+0.04	+1.2%	69%	36.45	0.58	0.67
	HMM	Conformer <sub>S</sub> <sup>FT</sup>	Orig. 800M	AED/CTC	Conformer <sub>S</sub> <sup>MW</sup>	-	+0.41	+11%	100%	39.24	0.48	0.52
	HMM	Conformer <sub>L</sub>	Transcript	AED/CTC	Conformer <sub>L</sub>	-	+0.18	+6.3%	99%	29.31	0.60	0.73
	HMM	Conformer <sub>L</sub>	Orig. 800M	AED/CTC	Conformer <sub>L</sub>	-	+0.21	+7.2%	99%	32.25	0.53	0.62
	HMM	Conformer <sub>L</sub>	Transcript	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	-0.03	-1.2%	67%	27.44	0.61	0.71
	HMM	Conformer <sub>L</sub>	Orig. 800M	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	-0.01	-0.2%	52%	30.95	0.53	0.61
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	-0.36	-14%	100%	29.20	0.62	0.70
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Orig. 800M	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	-0.27	-10%	100%	32.33	0.52	0.54
	AED/CTC	Conformer <sub>L</sub>	-	AED/CTC	Conformer <sub>S</sub>	-	+0.78	+21%	100%	31.76	0.64	0.70
	AED/CTC	w2v2.0	-	AED/CTC	Conformer <sub>L</sub>	-	+0.44	+16%	100%	30.92	0.59	0.71
	AED/CTC	w2v2.0	-	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	+0.23	+8.8%	100%	29.85	0.60	0.69
	AED/CTC	w2v2.0	-	AED/CTC	Conformer <sub>S</sub>	-	+1.22	+34%	100%	35.65	0.57	0.67
	AED/CTC	w2v2.0	-	AED/CTC	Conformer <sub>S</sub> <sup>MW</sup>	-	+1.20	+33%	100%	35.73	0.57	0.66
	AED/CTC	w2v2.0	-	AED	CRDNN	-	+2.70	+53%	100%	45.92	0.47	0.58
	AED/CTC	Conformer <sub>L</sub>	-	AED	CRDNN	-	+2.26	+44%	100%	43.21	0.51	0.58
	AED/CTC	Conformer <sub>S</sub>	-	AED	CRDNN	-	+1.48	+29%	100%	43.05	0.54	0.65
	HMM	Conformer <sub>L</sub>	Transcript	HMM	Conformer <sub>S</sub>	Transcript	+0.91	+25%	100%	32.06	0.60	0.70
	HMM	Conformer <sub>L</sub>	Orig. 800M	HMM	Conformer <sub>S</sub>	Orig. 800M	+0.70	+21%	100%	26.26	0.67	0.72
	HMM	w2v2.0	Transcript	HMM	Conformer <sub>L</sub>	Transcript	+0.43	+16%	100%	28.02	0.58	0.67
	HMM	w2v2.0	Transcript	HMM	Conformer <sub>S</sub>	Transcript	+1.34	+37%	100%	35.42	0.51	0.58
	HMM	w2v2.0	Transcript	HMM	CRDNN	Transcript	+2.32	+51%	100%	42.98	0.45	0.46
	HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	HMM	CRDNN	Transcript	+1.87	+41%	100%	40.92	0.51	0.62
	HMM	Conformer <sub>S</sub>	Transcript	HMM	CRDNN	Transcript	+0.98	+22%	100%	38.66	0.59	0.72
	HMM	CRDNN	Orig. 800M	HMM	CRDNN	Transcript	+0.67	+15%	100%	35.23	0.62	0.67
	HMM	CRDNN	Transcript	FS HMM	CRDNN	Transcript	+0.79	+15%	100%	43.44	0.58	0.75
	HMM	CRDNN	Orig. 800M	FS HMM	CRDNN	Orig. 800M	+0.52	+12%	100%	33.85	0.65	0.70
	HMM	Conformer <sub>L</sub>	Transcript	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	+0.33	+11%	100%	27.82	0.63	0.71
	HMM	Conformer <sub>L</sub>	Orig. 800M	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Orig. 800M	+0.27	+9.2%	100%	26.30	0.66	0.73
	FS HMM	CRDNN	Orig. 800M	FS HMM	CRDNN	Transcript	+0.94	+18%	100%	42.14	0.57	0.55
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Orig. 800M	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	+0.09	+3.0%	85%	25.34	0.63	0.58
	HMM	Conformer <sub>S</sub> <sup>FT</sup>	Transcript	HMM	Conformer <sub>S</sub>	Transcript	+0.02	+0.4%	69%	6.87	0.94	0.92
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	FS HMM	Conformer <sub>L</sub>	Transcript	+0.08	+2.7%	95%	17.44	0.80	0.86
	AED/CTC	Conformer <sub>S</sub> <sup>MW</sup>	-	AED/CTC	Conformer <sub>S</sub>	-	+0.02	+0.6%	75%	9.16	0.92	0.90
	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	AED/CTC	Conformer <sub>L</sub>	-	+0.21	+7.4%	100%	13.47	0.84	0.81

TABLE III

PAIRWISE DIFFERENCES OF SELECTED SYSTEMS. CREDIBILITY (CREDIB.) IS PROBABILITY OF IMPROVEMENT (OF THE BETTER SYSTEM) MEASURED BY BOOTSTRAP ESTIMATE [1]. SENTENCE DIFFERENCE (S-DIFF.) IS THE PERCENTAGE OF TRANSCRIPTIONS THAT WERE DIFFERENT BETWEEN A AND B. THE SUPERSCRIPTS  $^{MW}$  AND  $^{FT}$  MEAN THE POST-10-MODEL-AVERAGING MINIMUM WORD ERROR RATE FINETUNING AND HMM FINETUNING, RESPECTIVELY.

	A			B			WER Difference (B-A)			S-Diff.	WER Rank $\tau$	
	Family	Encoder	LM Data	Family	Encoder	LM Data	[pp.]	Relative	Credib.	[%]	Utt.	Spk.
Librispeech Test Other	HMM	CRDNN	Transcript	AED	CRDNN	-	+2.46	+16%	100%	75.88	0.53	0.80
	HMM	CRDNN	Orig. 800M	AED	CRDNN	-	+4.58	+30%	100%	76.62	0.48	0.74
	FS HMM	CRDNN	Transcript	AED	CRDNN	-	+1.35	+9.0%	100%	77.99	0.53	0.81
	FS HMM	CRDNN	Orig. 800M	AED	CRDNN	-	+3.93	+26%	100%	79.55	0.43	0.70
	HMM	w2v2.0	Transcript	AED/CTC	w2v2.0	-	+1.02	+17%	100%	47.29	0.60	0.78
	HMM	Conformer <sub>S</sub>	Transcript	AED/CTC	Conformer <sub>S</sub>	-	+0.61	+6.9%	100%	59.65	0.57	0.75
	HMM	Conformer <sub>S</sub>	Orig. 800M	AED/CTC	Conformer <sub>S</sub>	-	+2.08	+23%	100%	61.01	0.52	0.71
	HMM	Conformer <sub>S</sub> <sup>FT</sup>	Transcript	AED/CTC	Conformer <sub>S</sub> <sup>MW</sup>	-	+0.66	+7.4%	100%	59.24	0.58	0.78
	HMM	Conformer <sub>S</sub> <sup>FT</sup>	Orig. 800M	AED/CTC	Conformer <sub>S</sub> <sup>MW</sup>	-	+2.04	+23%	100%	61.18	0.52	0.71
	HMM	Conformer <sub>L</sub>	Transcript	AED/CTC	Conformer <sub>L</sub>	-	-0.17	-2.8%	94%	45.93	0.64	0.79
	HMM	Conformer <sub>L</sub>	Orig. 800M	AED/CTC	Conformer <sub>L</sub>	-	+0.32	+5.5%	100%	47.91	0.58	0.71
	HMM	Conformer <sub>L</sub>	Transcript	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	-0.08	-1.3%	77%	46.55	0.63	0.78
	HMM	Conformer <sub>L</sub>	Orig. 800M	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	+0.41	+6.8%	100%	48.79	0.57	0.72
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	AED/CTC	Conformer <sub>L</sub>	-	-0.69	-12%	100%	49.17	0.61	0.76
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Orig. 800M	AED/CTC	Conformer <sub>L</sub>	-	-0.02	-0.3%	55%	51.75	0.52	0.67
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	-0.60	-10%	100%	50.09	0.59	0.76
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Orig. 800M	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	+0.07	+1.2%	72%	52.67	0.51	0.67
	AED/CTC	Conformer <sub>L</sub>	-	AED/CTC	Conformer <sub>S</sub>	-	+2.95	+33%	100%	52.88	0.61	0.80
	AED/CTC	w2v2.0	-	AED/CTC	Conformer <sub>L</sub>	-	-0.13	-2.2%	59%	51.72	0.57	0.76
	AED/CTC	w2v2.0	-	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	-0.04	-0.7%	50%	51.82	0.57	0.74
	AED/CTC	w2v2.0	-	AED/CTC	Conformer <sub>S</sub>	-	+2.82	+32%	100%	60.67	0.54	0.74
	AED/CTC	w2v2.0	-	AED/CTC	Conformer <sub>S</sub> <sup>MW</sup>	-	+2.78	+31%	100%	60.97	0.53	0.75
	AED/CTC	w2v2.0	-	AED	CRDNN	-	+8.96	+60%	100%	74.48	0.43	0.65
	AED/CTC	Conformer <sub>L</sub>	-	AED	CRDNN	-	+2.26	+44%	100%	43.21	0.51	0.58
	AED/CTC	Conformer <sub>S</sub>	-	AED	CRDNN	-	+1.48	+29%	100%	43.05	0.54	0.65
	HMM	Conformer <sub>L</sub>	Transcript	HMM	Conformer <sub>S</sub>	Transcript	+2.17	+26%	100%	50.63	0.64	0.81
	HMM	Conformer <sub>L</sub>	Orig. 800M	HMM	Conformer <sub>S</sub>	Orig. 800M	+1.19	+18%	100%	44.27	0.66	0.81
	HMM	w2v2.0	Transcript	HMM	Conformer <sub>L</sub>	Transcript	+1.06	+17%	100%	49.17	0.56	0.76
	HMM	w2v2.0	Transcript	HMM	Conformer <sub>S</sub>	Transcript	+3.23	+39%	100%	57.23	0.53	0.73
	HMM	w2v2.0	Transcript	HMM	CRDNN	Transcript	+7.53	+60%	100%	69.45	0.43	0.66
	HMM	Conformer <sub>L</sub>	Transcript	HMM	CRDNN	Transcript	+6.46	+51%	100%	67.57	0.50	0.70
	HMM	Conformer <sub>S</sub> <sup>FT</sup>	Transcript	HMM	CRDNN	Transcript	+4.37	+35%	100%	65.09	0.55	0.69
	HMM	CRDNN	Orig. 800M	HMM	CRDNN	Transcript	+2.12	+17%	100%	62.61	0.63	0.81
	HMM	CRDNN	Transcript	FS HMM	CRDNN	Transcript	+1.11	+8.1%	100%	69.48	0.61	0.81
	HMM	CRDNN	Orig. 800M	FS HMM	CRDNN	Orig. 800M	+0.65	+5.9%	100%	62.78	0.60	0.81
	HMM	Conformer <sub>L</sub>	Transcript	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	+0.52	+7.9%	100%	46.51	0.63	0.79
	HMM	Conformer <sub>L</sub>	Orig. 800M	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Orig. 800M	+0.34	+5.7%	100%	43.11	0.62	0.81
	FS HMM	CRDNN	Orig. 800M	FS HMM	CRDNN	Transcript	+2.58	+19%	100%	68.49	0.58	0.74
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Orig. 800M	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	+0.68	+10%	100%	40.63	0.66	0.73
	HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	HMM	Conformer <sub>S</sub>	Transcript	+0.08	+1.0%	96%	16.40	0.91	0.93
	FS HMM	Conformer <sub>L</sub> <sup>FT</sup>	Transcript	FS HMM	Conformer <sub>L</sub>	Transcript	-0.00	-0.1%	52%	34.60	0.75	0.85
	AED/CTC	Conformer <sub>S</sub> <sup>MW</sup>	-	AED/CTC	Conformer <sub>S</sub>	-	+0.04	+0.5%	75%	21.61	0.89	0.93
	AED/CTC	Conformer <sub>L</sub> <sup>MW</sup>	-	AED/CTC	Conformer <sub>L</sub>	-	-0.09	-1.5%	94%	19.60	0.87	0.91