Context	RP 1	RP 2	RP 3	AVG	STD
5s	8.86 / 7.72	8.78 / 7.80	8.51 / 7.57	8.72 / 7.70	0.18 / 0.12
10s	7.69 / 7.15	8.14 / 7.15	8.23 / 7.38	8.02 / 7.23	0.29 / 0.13
20s	7.70 / 7.12	7.83 / 6.93	7.86 / 7.07	**7.80** / **7.04**	0.09 / 0.1
40s	8.02 / 7.53	8.08 / 7.53	7.94 / 7.58	8.01 / 7.55	0.07 / 0.03

Table 1: Single Utterance Tedlium (Dev/Test)

Context	RP 1	RP 2	RP 3	AVG	STD
5s	8.44 / 7.49	8.63 / 7.52	8.22 / 7.26	8.43 / 7.42	0.21 / 0.14
10s	7.69 / 6.72	7.88 / 6.74	7.84 / 7.0	7.80 / 6.82	0.10 / 0.16
20s	7.24 / 6.51	7.48 / 6.61	7.39 / 6.63	7.37 / 6.58	0.12 / 0.06
40s	7.14 / 6.51	7.17 / 6.42	7.34 / 6.66	**7.22** / **6.53**	0.11 / 0.12

Table 2: Tedlium 87.5% overlap constant sequence length

Context	RP 1	RP 2	RP 3	AVG	STD
5s	28.82 / 22.30	28.51 / 21.92	28.42/ 21.56	28.58 / 21.93	0.21 / 0.37
10s	27.07 / 20.72	26.91 / 20.41	26.99 / 20.45	26.99 / 20.53	0.08 / 0.17
20s	25.15 / 19.21	25.64 / 19.64	25.66 / 19.47	25.48 / 19.44	0.29 / 0.22
40s	25.19 / 19.19	25.02 / 19.35	25.40 / 19.52	**25.20** / **19.35**	0.19 / 0.17

Table 3: Earnings-22 87.5% overlap constant sequence length

References

Context	RP 1	RP 2	RP 3	AVG	STD
20s	26.47 / 20.34	25.38 / 19.44	25.17 / 19.08	25.67 / 19.62	0.70 / 0.65
40s	24.83 / 19.46	24.70 / 19.03	25.31 / 19.22	24.95 / 19.24	0.32 / 0.22
80s	24.29 / 18.70	24.42 / 18.62	24.44 / 18.70	24.38 / 18.67	0.08 / 0.05
160s	24.54 / 19.03	23.80 / 18.35	24.25 / 18.73	24.20 / 18.70	0.37 / 0.34
320s	24.02 / 18.54	24.02 / 18.59	24.29 / 18.70	**24.11** / **18.61**	0.16 / 0.08
3600s	24.49 / 19.0	24.30 / 18.56	24.24 / 18.89	24.34 / 18.81	0.13 / 0.23

Table 4: Earnings-22 87.5% overlap sequence length warmup

Context	RP 1	RP 2	RP 3	AVG	STD
20s	7.72 / 6.97	7.67 / 6.75	7.41 / 6.67	7.60 / 6.80	0.17 / 0.16
40s	7.40 / 6.54	7.49 / 6.71	7.46 / 6.69	7.45 / 6.65	0.05 / 0.09
80s	7.23 / 6.44	7.23 / 6.53	7.16 / 6.44	**7.21** / 6.47	0.04 / 0.05
160s	7.41 / 6.38	7.09 / 6.50	7.28 / 6.51	7.26 / 6.46	0.16/0.07
320s	7.21 / 6.50	7.27 / 6.52	7.30 / 6.33	7.26 / 6.45	0.05 / 0.10
3600s	7.42 / 6.48	7.20 / 6.39	7.0 / 6.40	**7.21** / **6.42**	0.21 / 0.05

Table 5: Tedlium 87.5% overlap sequence

Context	RP 1	RP 2	RP 3	AVG
5s	13.72/12.79	13.84/13.02	13.58/12.73	13.71/12.85
20s*	8.56/7.95	8.78/8.04	8.70/8.01	8.68/8.0
80s	7.54/6.79	7.57/6.90	7.38/6.78	7.50/6.82

Table 6: 0% Overlap. *no sequence length warmup

Context	RP 1	RP 2	RP 3	AVG
5s	10.95/10.20	11.6/10.40	11.24/10.24	11.26/10.28
20s*	7.88/7.26	8.30/7.39	8.17/7.47	8.12/7.38
80s	7.50/6.64	7.51/6.73	7.34/6.63	7.45/6.67

Table 7: 25% Overlap. *no sequence length warmup

Context	RP 1	RP 2	RP 3	AVG
5s	11.90/11.28	12.14/11.41	12.26/11.22	12.10/11.30
20s*	8.21/7.48	8.59/7.68	8.36/7.74	8.39/7.64
80s	7.54/6.71	7.55/6.85	7.39/6.70	7.50/6.75

Table 8: 50% Overlap. *no sequence length warmup

Context	RP 1	RP 2	RP 3	AVG
5s	8.56/7.59	8.76/7.64	8.26/7.43	8.53/7.56
20s*	7.26/6.61	7.46/6.67	7.34/6.67	7.35/6.65
80s	7.28/6.50	7.23/6.58	7.21/6.44	7.24/6.51

Table 9: 75% Overlap. *no sequence length warmup

Context	RP 1	RP 2	RP 3	AVG
5s	8.44/7.49	8.63/7.52	8.22/7.26	8.43/7.42
20s*	7.24/6.51	7.48/6.61	7.39/6.63	7.37/6.58
80s	7.23/6.44	7.23/6.53	7.16/6.44	7.21/6.47

Table 10: 87.5% Overlap. *no sequence length warmup

Context	RP 1	RP 2	RP 3	AVG
5s	8.36/7.49	8.57/7.51	8.16/7.21	8.36/7.41
20s*	7.21/6.46	7.49/6.59	7.49/6.63	7.37/6.56
80s	7.23/6.45	7.20/6.52	7.15/6.41	7.20/6.46

Table 11: 93.75% Overlap. *no sequence length warmup

		Evaluation					
	Context	2.5s	5s	10s	20s	40s	80s
Training	5s	-	-	-	-	-	-
	10s	-	-	-	-	-	-
	20s	-	-	-	-	-	-
	40s	-	-	-	-	-	-

Table 12

Dataset	64	128	256	512	1024
Tedlium	98.1	83.1	75.6	71.6	69.7
Earnings-22	83.4	69.4	62.3	57.6	54.0

Table 13: Perplexity (PPL) for TLMs at (,,,,) tokens of context

Dataset	AM Context	64	128	256	512	1024
Earnings-22	5s	23.14 / 17.48	22.94 / 17.37	22.79 / 17.26	22.79 / 17.19	22.80 / 17.12
	80s	19.48 / 14.98	19.14/ 14.74	19.05 / 14.65	19.0 / 14.63	18.92 / 14.57
Tedlium	5s	5.79 / 5.44	5.74 / 5.36	5.87 / 5.32	5.82 / 5.32	5.82 / 5.32
	80s	5.16 / 4.78	5.20 / 4.80	5.11 / 4.75	5.15 / 4.73	5.11 / 4.74

Table 14: WERs when decoding with LM via beam search for various AM (s) and LM context sizes (number of tokens). **repeat 1.** (parameters tuned for this repeat)

Dataset	Dev	Test
Earnings-22		13.64
Tedlium	4.85	4.44!

Table 15: REPEAT 1 - 2 EPOCHS AM CONTEXT: 80s - LM CONTEXT: 1024

Dataset	AM Context	64	128	256	512	1024
Earnings-22	5s	22.65 / 17.16	22.39 / 16.93	22.28 / 16.81	22.25 / 16.67	22.26 / 16.66
	80s	19.60 / 14.76	19.27 / 14.62	19.08 / 14.55	19.10 / 14.57	19.09 / 14.41
Tedlium	5s	6.05 / 5.39	6.01 / 5.31	5.94 / 5.29	6.02 / 5.35	5.98 / 5.36
	80s	5.07 / 4.87	4.95 / 4.81	4.94 / 4.81	5.05 / 4.81	4.99 / 4.83

Table 16: WERs when decoding with LM via beam search for various AM (s) and LM context sizes (number of tokens). **repeat 2.**

Dataset	AM Context	64	128	256	512	1024
Earnings-22	5s	22.76 / 16.73	22.46 / 16.62	22.33 / 16.53	22.31 / 16.49	22.28 / 16.40
	80s	19.64 / 14.74	19.45 / 14.60	19.32 / 14.49	19.36 / 14.50	19.23 / 14.44
Tedlium	5s	5.74 / 5.38	5.70 / 5.29	5.66 / 5.23	5.63 / 5.26	5.63 / 5.29
	80s	5.25 / 4.92	5.16 / 4.81	5.11 / 4.84	5.18 / 4.84	5.17 / 4.87

Table 17: WERs when decoding with LM via beam search for various AM (s) and LM context sizes (number of tokens). **repeat 3.**

Dataset	AM Context	64	128	256	512	1024
Earnings-22	5s	22.85 / 17.12	22.60 / 16.97	22.47 / 16.87	22.45 / 16.78	22.45 / 16.73
	80s	19.57 / 14.83	19.29 / 14.65	19.15 / 14.56	19.15 / 14.57	19.08 / 14.47
Tedlium	5s	5.86 / 5.40	5.82 / 5.32	5.82 / 5.28	5.82 / 5.31	5.81 / 5.32
	80s	5.16 / 4.86	5.10 / 4.81	5.05 / 4.8	5.13 / 4.79	5.09 / 4.81

Table 18: WERs when decoding with LM via beam search for various AM (s) and LM context sizes (number of tokens). **AVERAGE OF REPEATS**

Dataset	AM Context	64	128	256	512	1024
Earnings-22	5s					
	80s					
Tedlium	5s					
	80s					

Table 19: TEMPLATE