

OWSM v3.1:

Better and Faster Open Whisper-Style Speech Models based on E-Branchformer



HONDA

Yifan Peng¹, Jinchuan Tian¹, William Chen¹, Siddhant Arora¹, Brian Yan¹, Yui Sudo², Muhammad Shakeel², Kwanghee Choi¹, Jiatong Shi¹, Xuankai Chang¹, Jee-weon Jung¹, Shinji Watanabe¹

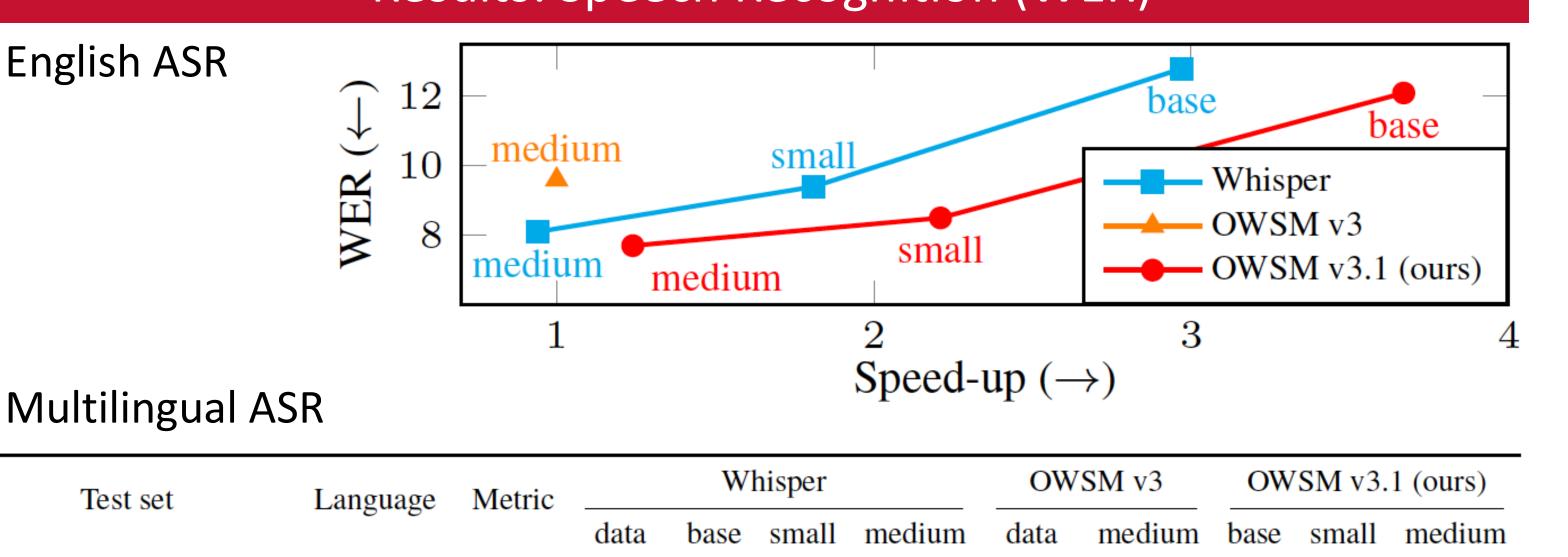
¹Carnegie Mellon University

²Honda Research Institute Japan

Introduction

- > OpenAl Whisper (Radford et al., 2023) has strong performance but does not release its complete training pipeline
- > Open Whisper-style Speech Model (OWSM) (Peng et al., 2023) is an initial step towards reproducing Whisper using publicly available data and open-source toolkits
- \succ This work proposes OWSM v3.1, the latest version of OWSM, which improves the performance and efficiency of previous versions with the same amount of training data
 - > OWSM v3.1 uses E-Branchformer as encoder

ESPnet



Results: Speech Recognition (WER)

 Three sizes: base (101M), small (367M), medium (1.02B) OWSM v3.1 outperforms previous OWSM in the following test sets while being 16% to 25% faster: 8 of 9 English ASR 10 of 11 multilingual ASR 13 of 19 ST 3 of 4 SLUE-PERB OWSM v3.1 shows emergent abilities in zero-shot contextual biasing 					MLS [24] AISHELL-1 [32] KsponSpeech clean [22] KsponSpeech other [22] ReazonSpeech [33] Average WER/CER (↓)	когеан	WER	11.1K 9.8K 13.3K 2.1K 2.6K 8.6K 4.3K 23.4K 8.0K 7.1K	$\begin{array}{rrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrrr$	$\begin{array}{cccccccccccccccccccccccccccccccccccc$	K 14.1 K 11.9 K 17.7 K 24.5 K 28.2 K 37.0 K 7.1 K 20.5 Z2.6	18.7 28.6 33.7 44.9 49.7 12.2 23.8	10.8 $9.$ 14.1 12 12.4 10 19.7 18 21.8 20 26.7 21 28.5 25 7.5 $6.$ 17.2 16 18.9 18 8.5 $7.$ 16.9 15
	Experi	imental Set	ups		(\v)						10.0		
OWSM v3.1 architecture					Long-form ASF	۲	Whisper		OWSM v3	OWSM v3.1 (ours		(ours)	
Encoder: E-Branchformer (OWSM v3 uses vanilla Transformer)					bas	ise si	mall medium		medium	base	small	medium	
Decoder: Trans		-				5.3	3 4	4.4	3.8	9.2	9.6	6.7	<u>5.7</u>
Training data amo	ount is the s	same as OWS	SM v3 (n	o new data)									
Model	Мо	del size Tra	ining da	ta (h) GPU hours		Results	: Spo	oken	Languag	ge Unders	standir	ng	
Whisper base		74M			The pre-traine	d speech	enco	der is	frozen, a	nd a rando	omly ini [.]	tialized	d shallov
Whisper small 244M		244M	680k	Unknown	decoder is trai	ned on ta	ask-sp	pecific	SLU data	from SLUI	E-PERB		
Whisper mediu		769M							. .				
OWSM v3		389M	180k	30.7k		Task		N	Aetric O	WSM v3	OWSM y	v3.1 (ou	urs)
OWSM v3.1 base				2 2L	Senti	ment Anal	ysis	F1	score	60.1	5	6.2	
	2C T	L01M		2.3k			•						
OWSM v3.1 sm		367M	180k		Named E	Entity Reco	U	on F1	score	54.8	6	5.8	
OWSM v3.1 sm	all 3	867M	180k	3.2k	Named E Named E	Entity Reco Entity Loca	lizati	on Fl on fra	score me-F1	54.8 40.5	6 5	5.8 0.4	
OWSM v3.1 sm OWSM v3.1 med	all 3 lium 10 w-	367M 020M		3.2k 24.6k	Named E Named E	Entity Reco	lizati	on Fl on fra	score me-F1	54.8	6 5	5.8	
OWSM v3.1 sm	all 3 lium 10 w-	867M	180k 70k	3.2k	Named E Named E	Entity Reco Entity Loca	lizati	on Fl on fra	score me-F1	54.8 40.5	6 5	5.8 0.4	
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov	all 3 lium 10 w-	367M 020M		3.2k 24.6k 3.2k	Named E Named E	Entity Reco Entity Loca Act Classi	lizati	on Fl on fra	score me-F1	54.8 40.5	6 5	5.8 0.4	
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov	all 3 lium 10 w-	367M 020M		3.2k 24.6k 3.2k	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU)	lizatio	on F1 on fra on F1	score me-F1 score	54.8 40.5 56.5	6 5 6	5.8 0.4 4.8	.1 (ours)
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov restriction	all 3 lium 10 w- 3	367M 020M 367M	70k	3.2k 24.6k 3.2k Results: Spee	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi	lizatio	on F1 on fra on F1	score me-F1	54.8 40.5 56.5	6 5 6	5.8 0.4 4.8	
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov restriction	all 3 lium 10 w- 3 Whisper	367M 020M 367M	70k /SM v3	3.2k24.6k3.2k3.2kResults: SpeeOWSM v3.1 (ours)	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe	dizatio ificati	on F1 on fra on F1	score me-F1 score g Data (h)	54.8 40.5 56.5 OWSM v3 medium	6 5 6 3 0 W base	5.8 0.4 4.8 /SM v3 small	mediu
OWSM v3.1 sma OWSM v3.1 med OWSM v3.1 lov restriction	all 3 lium 10 w- 3 Whisper base small n	B67M 020M B67M W nedium data	70k /SM v3 medium	3.2k24.6k3.2k3.2kResults: SpeeOWSM v3.1 (ours)base small medium	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU)	alizatio ificati et	on F1 on fra on F1 Trainin	score me-F1 score	54.8 40.5 56.5 OWSM v3	6 5 6	5.8 0.4 4.8 /SM v3	mediu 25.4
OWSM v3.1 sma OWSM v3.1 med OWSM v3.1 lov restriction to-English Source data b German 4.3K 1	all 3 lium 10 w- 3 Whisper base small n 1.4 25.0	867M 020M 867M 0W nedium data 33.6 0.2K	70k ZSM v3 medium 16.2	3.2k 24.6k 3.2k 3.2k Results: Spee OWSM v3.1 (ours) base small medium 7.3 15.1 17.1	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa	alizatio ificati et	on F1 on fra on F1 Trainin	score me-F1 score g Data (h)	54.8 40.5 56.5 OWSM v3 medium 25.4	6 5 6 8 0 8 0 8 5 6	5.8 0.4 4.8 /SM v3 small 22.8	mediu 25.4 19.6
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov restriction •to-English Source data b German 4.3K 1 Spanish 6.7K 1	all 3 ium 10 N- 3 Whisper 3 Dase small 11.4 25.0 19.2 32.8	867M 020M 867M 67M wedium data a33.6 0.2K 39.7 0.1K	70k 70k 20.5	3.2k 24.6k 3.2k Secure 2 OWSM v3.1 (ours) base 7.3 15.1 17.1 10.0 19.3 22.3	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa Catala Chine Persia	an an an se an	on F1 on fra on F1 Trainin	score me-F1 score g Data (h) 4.0K 4.0K 3.7K 3.7K 3.7K 3.8K	54.8 40.5 56.5 OWSM v3 medium 25.4 20.0 33.4 9.5	6 5 6 8 OW base 14.6 7.7 14.5 3.0	5.8 0.4 4.8 /SM v3 small 22.8 15.9 26.7 7.7	mediu 25.4 19.6 32.1 10.1
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov restriction to-English Source data b German 4.3K 1 Spanish 6.7K 1 French 4.5K 1	all 3 lium 10 w- 3 Whisper 3 base small 11.4 25.0 19.2 32.8 13.1 26.4	867M 020M 867M 0W nedium data 33.6 0.2K 39.7 0.1K 34.4 0.3K	70k 70k 20.5 21.7	$\begin{array}{c c} 3.2k \\ 24.6k \\ 3.2k \\ \hline Scale{1}\\ \hline Scale{1}\\$	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa Catala Chines Persia Estoni	an an an an an an	on F1 on fra on F1 Trainin 14 0 13 0 0 0	score me-F1 score g Data (h) 4.0K 4.0K 3.7K 3.7K 3.7K 3.7K 3.7K 3.7K	54.8 40.5 56.5 0WSM v3 medium 25.4 20.0 33.4 9.5 7.8	6 5 6 8 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0	5.8 0.4 4.8 /SM v3 /SM v3 small 22.8 15.9 26.7 7.7 5.8	mediu 25.4 19.6 32.1 <u>10.1</u> 7.7
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov restriction to-English Source data b German 4.3K 1 Spanish 6.7K 1 French 4.5K 1	all 3 ium 10 N- 3 Whisper 3 Dase small 11.4 25.0 19.2 32.8	867M 020M 867M 67M wedium data a33.6 0.2K 39.7 0.1K	70k 70k 20.5	3.2k 24.6k 3.2k Secure 2 OWSM v3.1 (ours) base 7.3 15.1 17.1 10.0 19.3 22.3	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa Catala Chines Persia Estoni Mongol	an an an an an lian	on F1 on fra on F1 Trainin	score me-F1 score g Data (h) 4.0K 4.0K 4.0K 3.7K 3.7K 3.7K 3.7K 3.7K 3.4K	54.8 40.5 56.5 56.5 OWSM v3 medium 25.4 20.0 33.4 9.5 7.8 3.1	6 5 6 8 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0	5.8 0.4 4.8 /SM v3 small 22.8 15.9 26.7 7.7 5.8 3.3	mediu 25.4 19.6 32.1 10.1 7.7 4.6
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov restriction to-English Source data b German 4.3K 1 Spanish 6.7K 1 French 4.5K 1	all 3 ium 10 N- 3 Whisper 3 base small 11.4 25.0 19.2 32.8 13.1 26.4 9.7 21.7	867M 020M 867M 0W nedium data 33.6 0.2K 39.7 0.1K 34.4 0.3K	70k 70k 20.5 21.7	$\begin{array}{c c} 3.2k \\ 24.6k \\ 3.2k \\ \hline Scale{1}\\ \hline Scale{1}\\$	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa Catala Chines Persia Estoni Mongol Turkis	an an an an an an an an an an an an an a	on F1 on fra on F1 Trainin	score me-F1 score g Data (h) 4.0K 4.0K 4.0K 3.7K 0.4K 0.4K 0.4K 0.4K 0.4K	54.8 40.5 56.5 56.5 OWSM v3 medium 25.4 20.0 33.4 9.5 7.8 3.1 6.1	6 5 6 8 0 8 0 8 0 8 14.6 7.7 14.5 3.0 1.8 1.0 1.2	5.8 0.4 4.8 /SM v3 /SM v3 small 22.8 15.9 26.7 7.7 5.8 3.3 4.8	mediu 25.4 19.6 32.1 <u>10.1</u> 7.7
OWSM v3.1 sma OWSM v3.1 med OWSM v3.1 lov restriction to-English Source data b German 4.3K 1 Spanish 6.7K 1 French 4.5K 1 Catalan 0.2K 9	all 3 ium 10 N- 3 Whisper 3 base small 11.4 25.0 19.2 32.8 13.1 26.4 9.7 21.7	367M 020M 367M 367M 00W 00W nedium data 33.6 0.2K 39.7 0.1K 34.4 0.3K 29.2 0.1K 34.2 -	70k 70k 70k 70k 70k 70k 70k 70k 70k 70k	$\begin{array}{c} 3.2k \\ 24.6k \\ \hline 3.2k \\ \hline 3.2k \\ \hline \\ $	Named E Named E Dialogue	Entity Reco Intity Loca Act Classi BLEU) Targe Germa Catala Chines Persia Estoni Mongol Turkis Arabi	an an an an an an an an an an an an an a	on F1 on fra on F1 Trainin	score me-F1 score g Data (h) 4.0K 4.0K 4.0K 3.7K 3.7K 3.7K 3.7K 3.7K 3.7K 3.7K 3.7	54.8 40.5 56.5 56.5 0WSM v3 medium 25.4 20.0 33.4 9.5 7.8 3.1 6.1 6.1 6.6	6 5 6 8 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0	5.8 0.4 4.8 /SM v3 /SM v3 small 22.8 15.9 26.7 7.7 5.8 3.3 4.8 3.3 4.8 5.1	mediu 25.4 19.6 32.1 10.1 7.7 <u>4.6</u> <u>6.5</u> <u>7.2</u>
OWSM v3.1 sma OWSM v3.1 med OWSM v3.1 lov restriction to-English Source data b German 4.3K 1 Spanish 6.7K 1 French 4.5K 1 Catalan 0.2K 9	all 3 ium 10 w- 3 Whisper 3 base small 11.4 25.0 19.2 32.8 13.1 26.4 9.7 21.7 13.4 26.5	367M 020M 367M 367M 00W 00W nedium data 33.6 0.2K 39.7 0.1K 34.4 0.3K 29.2 0.1K 34.2 -	70k 70k 70k 70k 70k 70k 70k 70k 70k 70k	3.2k 24.6k 3.2k 3.2k Results: Spee OWSM v3.1 (ours) base small medium 7.3 15.1 17.1 10.0 19.3 22.3 11.1 20.3 22.7 9.0 16.2 18.4 9.4 17.7 20.1	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa Catala Chines Persia Estoni Mongol Turkis Arabi Swedi	an an an an an an an an an an an an an a	on F1 on fra on F1 Trainin 14 0 13 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	score me-F1 score g Data (h) 4.0K 4.0K 4.0K 3.7K 0.4K 0.4K 0.4K 0.9K 0.9K 0.9K 0.9K 0.9K	54.8 40.5 56.5 56.5 OWSM v3 medium 25.4 20.0 33.4 9.5 7.8 3.1 6.1 6.1 6.1 6.1 6.1 6.6 19.9	6 5 6 8 0 8 0 8 14.6 7.7 14.5 3.0 1.8 1.0 1.2 1.6 8.1	5.8 0.4 4.8 /SM v3 /SM v3 small 22.8 15.9 26.7 7.7 5.8 3.3 4.8 5.1 16.6	mediu 25.4 19.6 32.1 10.1 7.7 4.6 6.5 7.2 20.3
OWSM v3.1 sma OWSM v3.1 med OWSM v3.1 lov restriction to-English Source	all 3 lium 10 N- 3 Whisper 3 Dase small 11.4 25.0 19.2 32.8 13.1 26.4 9.7 21.7 13.4 26.5 .14x 1.80x	367M 020M 367M 367M 67M 67M 00W nedium data 33.6 0.2K 39.7 0.1K 34.4 0.3K 29.2 0.1K 34.2 - 0.98x -	70k 70k 70k 70k 70k 70k 70k 70k 70k 70k	3.2k 24.6k 3.2k 3.2k Results: Spee OWS v3.1 (ours) base small medium 7.3 15.1 17.1 10.0 19.3 22.3 11.1 20.3 22.7 9.0 16.2 18.4 9.4 17.7 20.1 3.23x 2.26x 1.16x	Named E Named E Dialogue	Entity Reco Intity Loca Act Classi BLEU) Targe Germa Catala Chines Persia Estoni Mongol Turkis Arabi	an an an an an lian se an an an an an an an an an an an an an	on F1 on fra on F1 Trainin 14 0 13 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	score me-F1 score g Data (h) 4.0K 4.0K 4.0K 3.7K 3.7K 3.7K 3.7K 3.7K 3.7K 3.7K 3.7	54.8 40.5 56.5 56.5 0WSM v3 medium 25.4 20.0 33.4 9.5 7.8 3.1 6.1 6.1 6.6	6 5 6 8 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0 9 0	5.8 0.4 4.8 /SM v3 /SM v3 small 22.8 15.9 26.7 7.7 5.8 3.3 4.8 3.3 4.8 5.1	mediu 25.4 19.6 32.1 10.1 7.7 4.6 6.5 7.2 20.3 6.4
OWSM v3.1 sm OWSM v3.1 med OWSM v3.1 lov restriction to-English Source data b German 4.3K 1 Spanish 6.7K 1 Spanish 6.7K 1 French 4.5K 1 Catalan 0.2K 9 Ave. BLEU (†) 1 Speed-up (†) 2.	all 3 lium 10 w- 3 Whisper 3 base small 11.4 25.0 19.2 32.8 13.1 26.4 9.7 21.7 13.4 26.5 .14x 1.80x	867M 020M 867M 867M 1000 1000 1000 1000 1000 1000 1000 10	70k 70k 70k 70k 70k 70k 70k 70k 70k 70k	3.2k 24.6k 3.2k 3.2k 3.2k Results: Spee OWSM v3.1 (ours) base small 7.3 15.1 17.1 10.0 19.3 22.3 11.1 20.3 22.7 9.0 16.2 18.4 9.4 17.7 20.1 3.23x 2.26x 1.16x	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa Catala Chines Persia Estoni Mongol Turkis Arabi Swedi Latvia	an an an an an lian se an an lian sh ic sh an lian sh ic sh an lian	on F1 on fra on F1 Trainin 14 0 13 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	score me-F1 score g Data (h) 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K	54.8 40.5 56.5 56.5 OWSM v3 medium 25.4 20.0 33.4 9.5 7.8 3.1 6.1 6.1 6.1 6.6 19.9 6.3	6 5 6 8 0 8 0 8 14.6 7.7 14.5 3.0 1.8 1.0 1.2 1.6 8.1 1.3	5.8 0.4 4.8 /SM v3 /SM v3 small 22.8 15.9 26.7 7.7 5.8 3.3 4.8 5.1 16.6 4.4	mediu 25.4 19.6 32.1 10.1 7.7 4.6 6.5 7.2 20.3
OWSM v3.1 sma OWSM v3.1 med OWSM v3.1 lov restriction to-English Source data b German 4.3K 1 Spanish 6.7K 1 Spanish 6.7K 1 French 4.5K 1 Catalan 0.2K 9 Ave. BLEU (†) 1 Speed-up (†) 2.	all 3 ium 10 w- 3 w- 3 Whisper 3 base small 11.4 25.0 19.2 32.8 13.1 26.4 9.7 21.7 13.4 26.5 .14x 1.80x Abilities in per, OWSM	367M 020M 367M 367M 367M add ata 33.6 0.2K 39.7 0.1K 34.4 0.3K 29.2 0.1K 34.2 0.98x - Zero-Shot optionally ta	70k 70k 70k 70k 70k 70k 70k 70k 70k 70k	3.2k 24.6k 3.2k 3.2k Results: Spee OWS v3.1 (ours) base small medium 7.3 15.1 17.1 10.0 19.3 22.3 11.1 20.3 22.7 9.0 16.2 18.4 9.4 17.7 20.1 3.23x 2.26x 1.16x	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa Catala Chine Persia Estoni Mongol Turkis Arabi Swedi Latvia Sloveni	an an an an lian lian sh ic sh an lian lian lian lian lian l	on F1 on fra on F1 Trainin 14 0 13 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	score me-F1 score g Data (h) 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K	54.8 40.5 56.5 56.5 OWSM v3 medium 25.4 20.0 33.4 9.5 7.8 3.1 6.1 6.1 6.6 19.9 6.3 8.6	6 5 6 6 7 7 14.6 7.7 14.5 3.0 1.8 1.0 1.2 1.6 8.1 1.3 0.7	5.8 0.4 4.8 7SM v3 5mall 22.8 15.9 26.7 7.7 5.8 3.3 4.8 5.1 16.6 4.4 5.1 16.6 4.4 5.7	25.4 19.6 32.1 10.1 7.7 <u>4.6</u> <u>6.5</u> <u>7.2</u> <u>20.3</u> <u>6.4</u> <u>9.0</u>
OWSM v3.1 sma OWSM v3.1 med OWSM v3.1 low restriction to-English Source	all 3 ium 10 N- 3 Whisper 3 Dase small 1 11.4 25.0 1 19.2 32.8 3 13.1 26.4 3 9.7 21.7 3 13.4 26.5 3 13.4 26.5 3 13.4 26.5 3 13.4 26.5 3 14x 1.80x 3 Abilities in generation 3	367M 020M 367M 367M 67M a67M own a67M a67M own a67M a67M own a67M a7M a7M a7M a7M a7M a7M	70k 70k 20.5 20.5 21.7 16.8 18.8 1.00x Contex akes a tex	3.2k 24.6k 3.2k 3.2k 3.2k Results: Spee OWSM v3.1 (ours) base small 7.3 15.1 17.1 10.0 19.3 22.3 11.1 20.3 22.7 9.0 16.2 18.4 9.4 17.7 20.1 3.23x 2.26x 1.16x	Named E Named E Dialogue	Entity Reco Entity Loca Act Classi BLEU) Targe Germa Catala Chine Persia Estoni Mongol Turkis Arabi Swedi Latvia Sloven Tami	an an an an an lian se an an lian sh ic sh an lian lian sh ic sh an lian sh ic sh an lian se	on F1 on fra on F1 Trainin 14 0 13 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	score me-F1 score g Data (h) 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K 4.0K	54.8 40.5 56.5 56.5 OWSM v3 medium 25.4 20.0 33.4 9.5 7.8 3.1 6.1 6.1 6.1 6.1 6.1 6.1 6.3 8.6 0.0	6 5 6 8 0 9 0 9 14.6 7.7 14.5 3.0 1.8 1.0 1.2 1.6 8.1 1.3 0.7 0.0	5.8 0.4 4.8 /SM v3 /SM v3 small 22.8 15.9 26.7 7.7 5.8 3.3 4.8 5.1 16.6 4.4 5.1 16.6 4.4 5.7 0.0	mediu 25.4 19.6 32.1 10.1 7.7 4.6 6.5 7.2 20.3 6.4 9.0 0.0

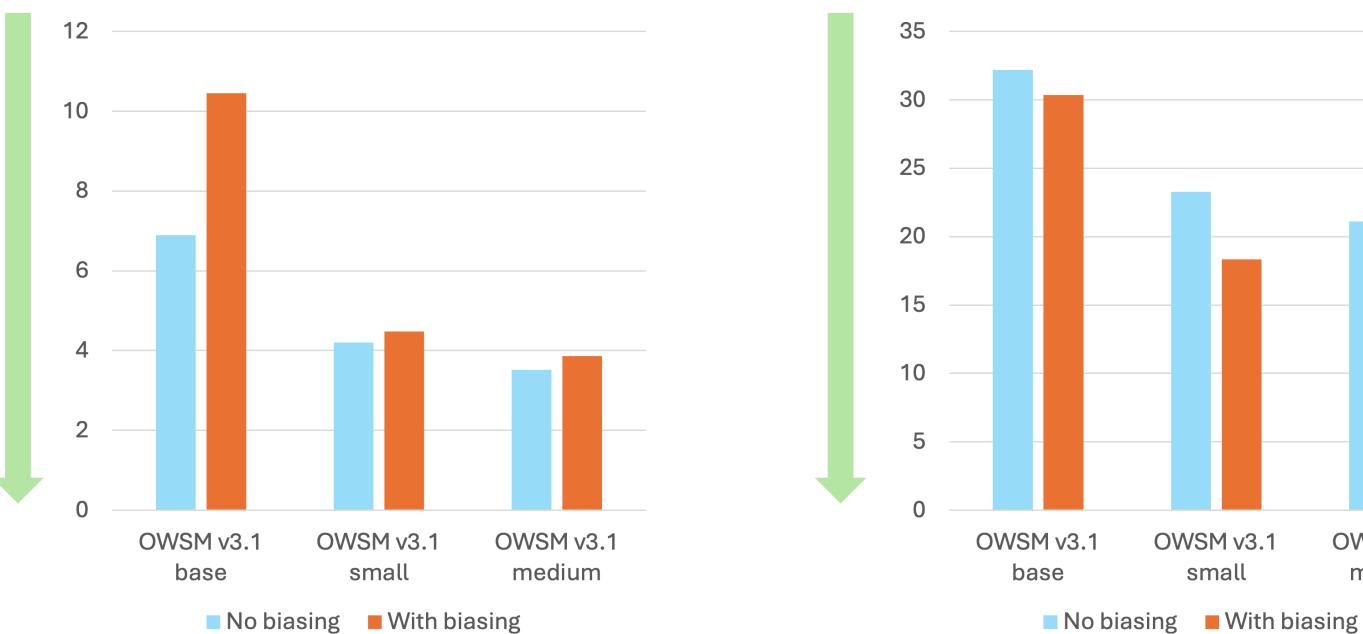
OWSM v3.1

medium

small

- > During inference, the user can provide a prompt to bias the output
- > We evaluate OWSM v3.1 on LibriSpeech biasing test sets (Le et al., 2021) > Contextual biasing aims to reduce the biased WER (B-WER) while maintaining the unbiased WER (U-WER) Biased WER (B-WER)

Unbiased WER (U-WER)





The phenomenon that the smaller OWSM performs poorly in zeroshot biasing ASR while larger ones perform well reveals that speech foundation models also have the emergent ability

References

- 1. Radford, Alec, et al., "Robust speech recognition via large-scale weak supervision," in Proc. ICML, 2023.
- 2. Peng, Yifan, et al., "Reproducing Whisper-Style Training Using An Open-Source Toolkit and Publicly Available Data," in Proc. ASRU, 2023.
- 3. Le, Duc, et al., "Contextualized streaming end-to-end speech recognition with trie-based deep biasing and shallow fusion," in Proc. Interspeech, 2021.