Table 6. Results on (a) English automatic speech recognition and (b) SUPERB benchmark. The WAV2VEC-960R is the baseline model trained on 960-hour LibriSpeech. The WAV2VEC-DIFFS4L is trained on the generated 960+960+480 data composition. The bolded results show better performance between the two models.

	(A) ENGLISH ASR		(B) SUPERB								
			KS	IC	SID	ER	QBE	SF		ASV	SD
TASK/METRIC	CER↓	WER↓	ACC↑	ACC↑	ACC↑	ACC↑	MTWV↑	F1↑	CER↓	EER↓	DER↓
WAV2VEC-960R	3.18	10.49	**96.23**	92.35	**75.18**	63.43	0.0233	88.30	24.77	**6.02**	6.08
WAV2VEC-DIFFS4L	**2.98**	**9.93**	96.17	**94.73**	65.79	**64.29**	**0.0630**	**88.50**	**24.71**	**6.02**	**6.03**