Robustness of pre-trained speech models

Kuan Po, Huang
Hung-yi, Lee

National Taiwan University

Publication

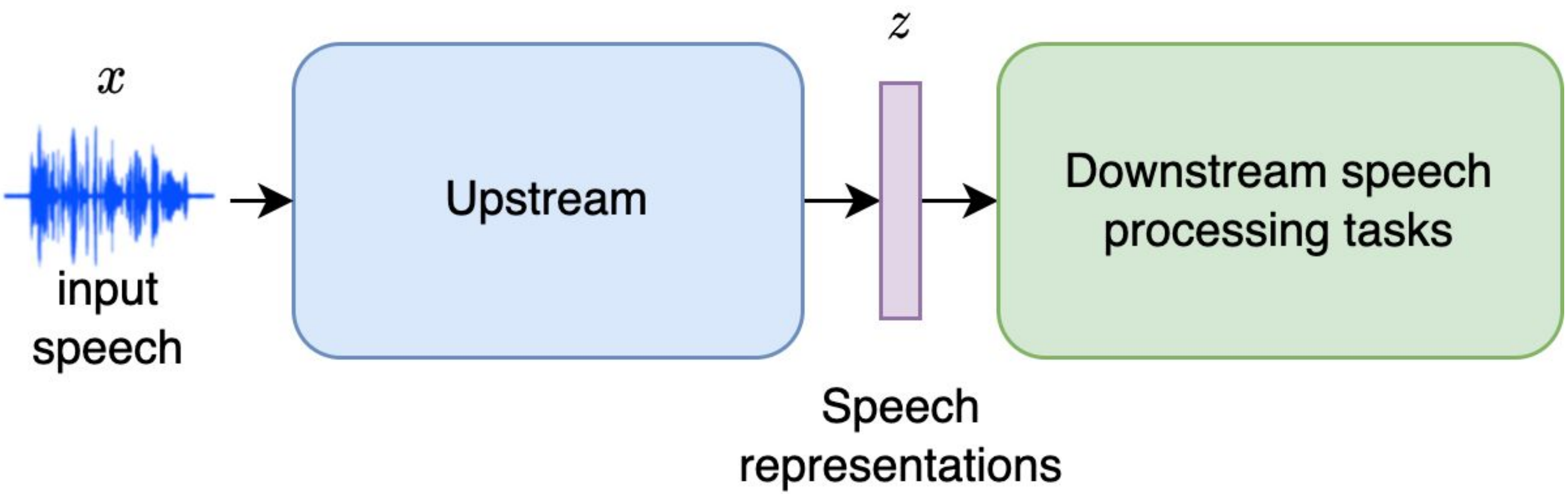
- Kuan Po Huang, Yu-Kuan Fu, Yu Zhang, and Hung-yi Lee. “**Improving Distortion Robustness of Self-supervised Speech Processing Tasks with Domain Adaptation**”. In: arXiv preprint arXiv:2203.16104 (2022).
- Submitted to Interspeech 2022.
- https://arxiv.org/abs/2203.16104

Improving Distortion Robustness of Self-supervised Speech Processing Tasks with Domain Adaptation

Kuan Po Huang[1] *Yu-Kuan Fu*[2] *Yu Zhang*[3] *Hung-yi Lee*[4]

[1]Graduate Institute of Computer Science and Information Engineering, National Taiwan University
[2]Department of Physics, National Taiwan University
[3]Google Brain
[4]Graduate Institute of Communication Engineering, National Taiwan University
[124]{r09922005, b07202024, hungyilee}@ntu.edu.tw, [3]ngyuzh@google.com

SUPERB



x
input
speech
Upstream
z
Downstream speech
processing tasks
Speech
representations

Domain mismatch

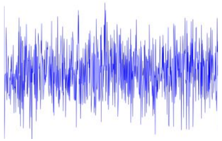
The training data and testing data have different distributions.

training data
(clean)

WER 6.72

testing data
(with distortions)

+

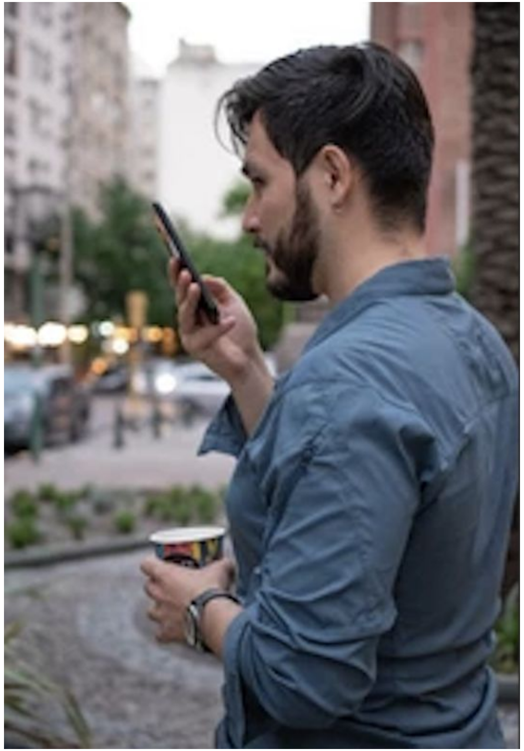


distortions

WER 10.16

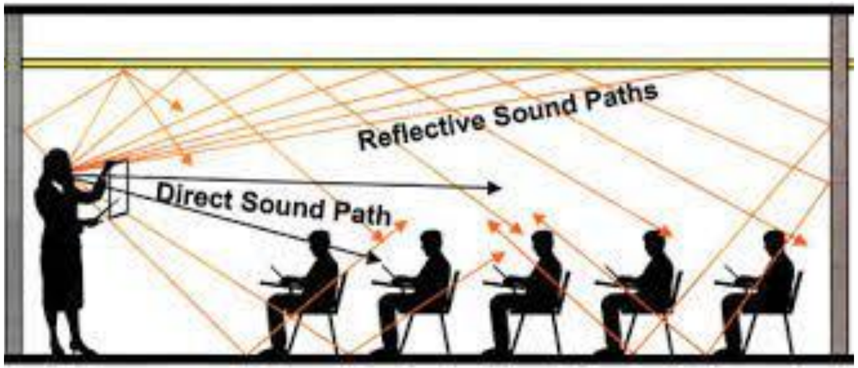
results from https://arxiv.org/abs/2203.16104

Recording environments

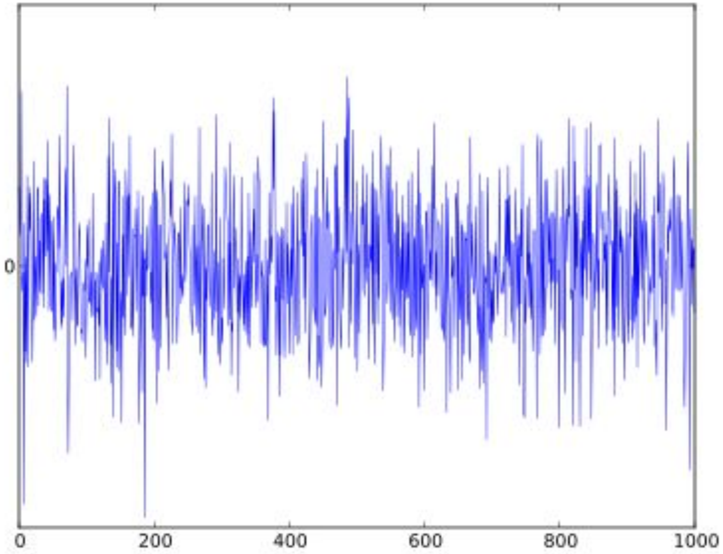


Speech distortions

reverberation

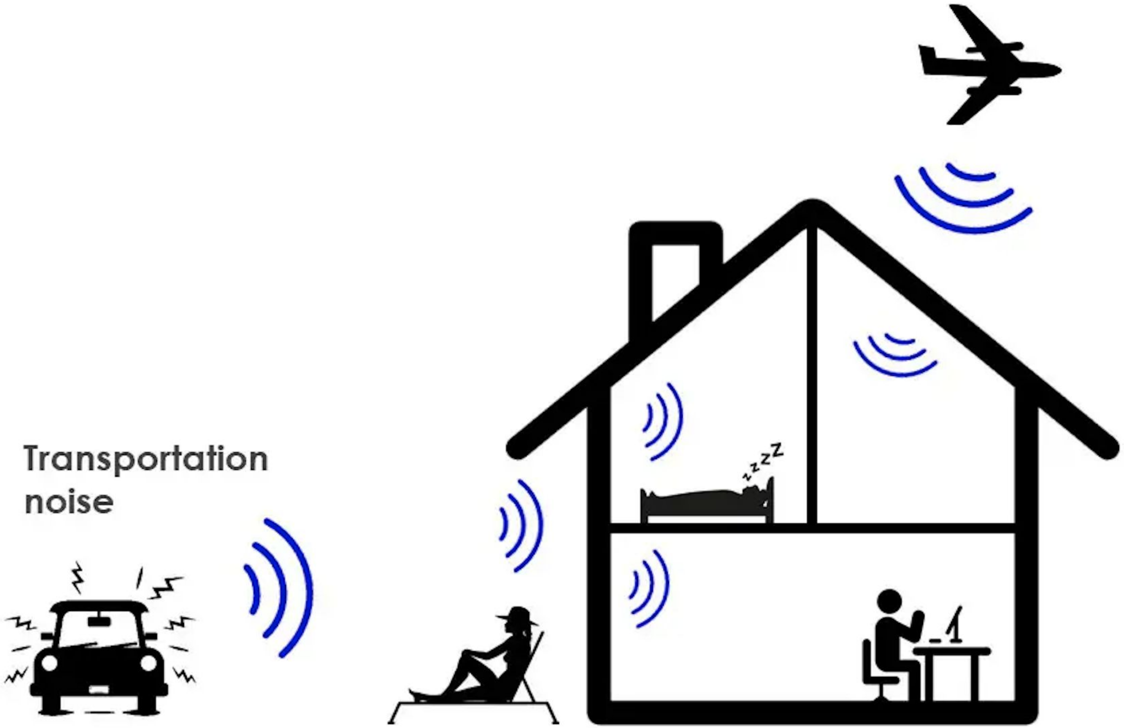


white noise



Speech distortions

- domestic noise
- transportation noise



Domain mismatch

Performance of SUPERB baselines in different domains of speech.

m: Musan noise, **g**: Gaussian noise, **r**: Reverberation

IC (Acc)			**ER** (Acc)			**KS** (Acc)		
clean	m+g+r	fsd50k	clean	m+g+r	fsd50k	clean	m+g+r	fsd50k
99.47	96.94	97.47	63.96	57.33	60.55	97.14	93.38	93.80

SID (Acc)			**ASR** (WER)							
			clean		m+g+r		fsd50k		CHiME3	
clean	m+g+r	fsd50k	w/o	w/ LM	w/o	w/ LM	w/o	w/ LM	w/o	w/ LM
84.97	65.51	77.61	6.72	4.88	10.16	7.94	9.62	7.57	33.4	29.26

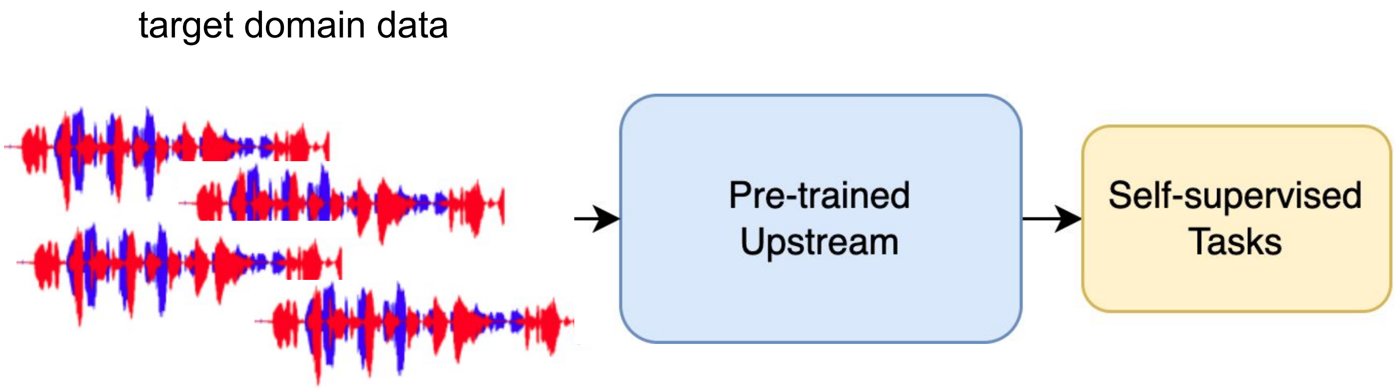
Domain adaptation

Large amount of unlabeled target domain data.

- **Upstream continual training**
- **Domain adversarial training**

Domain adaptation

Upstream continual training



Upstream continual training results

Performance of SUPERB baselines with upstream continually trained.

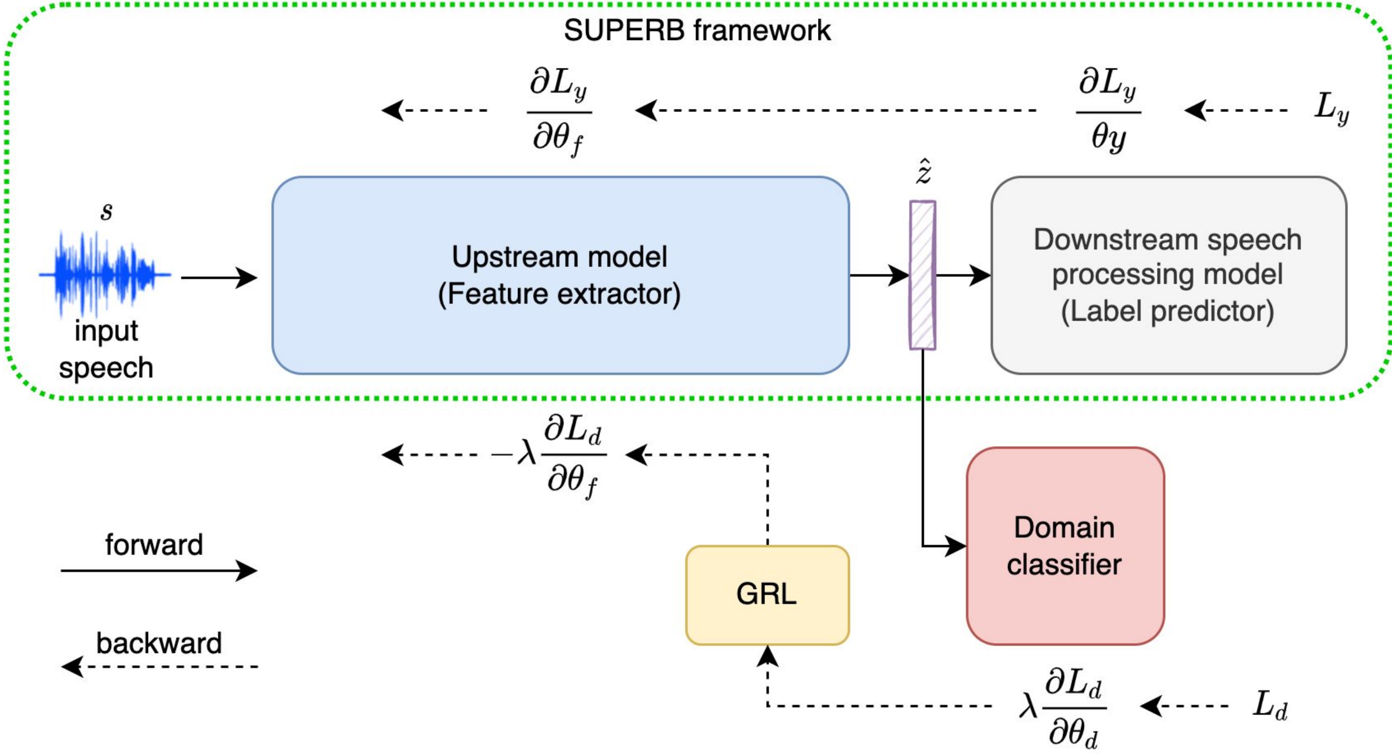
m: Musan noise, **g**: Gaussian noise, **r**: Reverberation

	continual	IC (Acc) clean	IC (Acc) m+g+r	IC (Acc) fsd50k	ER (Acc) clean	ER (Acc) m+g+r	ER (Acc) fsd50k	KS (Acc) clean	KS (Acc) m+g+r	KS (Acc) fsd50k
baseline	-	99.47	96.94	97.47	63.96	57.33	60.55	97.14	93.38	93.80
oracle	-	99.55	99.34	98.21	70.41	69.31	69.31	97.57	96.46	95.26
w/o DAT	libri 100hr mgr	99.45	98.63	97.94	64.42	62.30	60.65	96.92	94.87	93.90
w/o DAT	libri 960hr mgr	99.39	98.84	97.89	67.28	67.47	65.62	97.12	96.11	94.77

	continual	SID (Acc) clean	SID (Acc) m+g+r	SID (Acc) fsd50k	ASR (WER) clean w/o	ASR (WER) clean w/ LM	ASR (WER) m+g+r w/o	ASR (WER) m+g+r w/ LM	ASR (WER) fsd50k w/o	ASR (WER) fsd50k w/ LM	ASR (WER) CHiME3 w/o	ASR (WER) CHiME3 w/ LM
baseline	-	84.97	65.51	77.61	6.72	4.88	10.16	7.94	9.62	7.57	33.4	29.26
oracle	-	86.63	80.05	82.74	5.17	4.18	6.57	5.37	6.69	5.45	22.98	20.57
w/o DAT	libri 100hr mgr	87.02	70.91	80.96	6.23	4.87	8.04	6.47	7.90	6.38	27.82	24.27
w/o DAT	libri 960hr mgr	86.40	74.46	81.47	5.92	4.84	7.19	6.00	7.15	5.87	23.83	20.81

Domain adaptation

Domain adversarial training

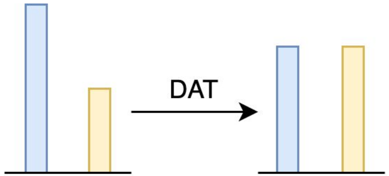


Binary / Multi-domain setting

Binary-domain

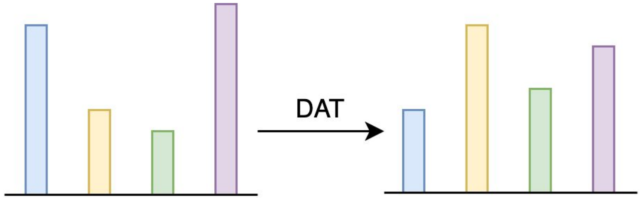
Binary cross entropy loss:

treat all distorted speech as the same domain

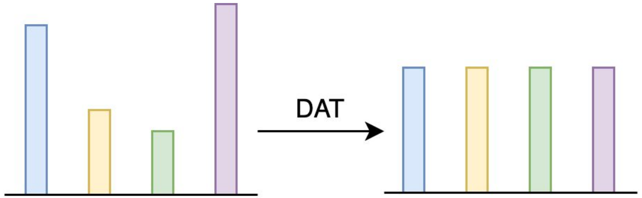


Multi-domain

Cross entropy loss



Entropy loss: uniform probability distribution



Domain adversarial training results

	continual	IC (Acc) clean	IC (Acc) m+g+r	IC (Acc) fsd50k	ER (Acc) clean	ER (Acc) m+g+r	ER (Acc) fsd50k	KS (Acc) clean	KS (Acc) m+g+r	KS (Acc) fsd50k
baseline	-	99.47	96.94	97.47	63.96	57.33	60.55	97.14	93.38	93.80
oracle	-	99.55	99.34	98.21	70.41	69.31	69.31	97.57	96.46	95.26
Cross Entropy Loss (CE)										
DAT $\lambda = 0.01$	-	99.47	98.68	97.47	68.85	63.59	63.50	**97.44**	95.26	94.64
DAT $\lambda = 0.001$	-	99.60	98.60	97.63	69.10	65.90	64.29	97.24	95.65	94.55
Entropy Loss (E)										
DAT $\lambda = 0.01$	-	99.55	98.47	97.36	63.87	59.91	59.26	96.92	94.94	94.06
DAT $\lambda = 0.001$	-	99.58	98.27	97.52	64.15	61.75	59.54	97.05	94.87	94.13
Binary Cross Entropy Loss (BCE)										
DAT $\lambda = 0.01$	-	**99.68**	98.39	97.57	66.18	64.33	62.86	96.98	95.10	93.93
DAT $\lambda = 0.001$	-	99.60	98.97	97.89	68.76	64.52	64.52	97.27	95.59	93.90

Domain adversarial training results

	continual	SID (Acc) clean	SID (Acc) m+g+r	SID (Acc) fsd50k	ASR (WER) clean w/o	ASR (WER) clean w/ LM	ASR (WER) m+g+r w/o	ASR (WER) m+g+r w/ LM	ASR (WER) fsd50k w/o	ASR (WER) fsd50k w/ LM	ASR (WER) CHiME3 w/o	ASR (WER) CHiME3 w/ LM
baseline	-	84.97	65.51	77.61	6.72	4.88	10.16	7.94	9.62	7.57	33.4	29.26
oracle	-	86.63	80.05	82.74	5.17	4.18	6.57	5.37	6.69	5.45	22.98	20.57
Cross Entropy Loss (CE)												
DAT $\lambda = 0.01$	-	86.49	71.82	79.76	6.16	4.60	11.74	9.97	11.39	12.27	44.43	41.51
DAT $\lambda = 0.001$	-	87.44	72.28	79.94	6.29	5.77	9.70	9.00	9.67	8.68	32.98	29.03
Entropy Loss (E)												
DAT $\lambda = 0.01$	-	90.01	75.23	84.04	5.59	**4.45**	11.79	10.23	8.64	7.25	40.02	37.45
DAT $\lambda = 0.001$	-	90.50	73.43	84.46	6.01	4.59	9.97	7.74	8.88	7.31	32.77	28.90
Binary Cross Entropy Loss (BCE)												
DAT $\lambda = 0.01$	-	89.25	73.71	82.12	6.23	4.68	9.68	7.56	9.18	7.38	31.20	28.72
DAT $\lambda = 0.001$	-	**90.96**	78.69	84.67	6.35	4.72	9.45	7.34	9.11	7.38	32.25	27.88

Domain adversarial training with continual training

	continual	IC (Acc) clean	IC (Acc) m+g+r	IC (Acc) fsd50k	ER (Acc) clean	ER (Acc) m+g+r	ER (Acc) fsd50k	KS (Acc) clean	KS (Acc) m+g+r	KS (Acc) fsd50k
(a) baseline	-	99.47	96.94	97.47	63.96	57.33	60.55	97.14	93.38	93.80
(b) oracle	-	99.55	99.34	98.21	70.41	69.31	69.31	97.57	96.46	95.26
(c) w/o DAT	libri 100hr mgr	99.45	98.63	97.94	64.42	62.30	60.65	96.92	94.87	93.90
(d) w/o DAT	libri 960hr mgr	99.39	98.84	97.89	67.28	67.47	65.62	97.12	96.11	94.77
					Cross Entropy Loss (CE)					
(e) DAT $\lambda = 0.01$	-	99.47	98.68	97.47	68.85	63.59	63.50	**97.44**	95.26	94.64
(f) DAT $\lambda = 0.001$	-	99.60	98.60	97.63	69.10	65.90	64.29	97.24	95.65	94.55
					Cross Entropy Loss (CE)					
(g) DAT $\lambda = 0.001$	libri 100hr mgr	99.66	**99.45**	**98.55**	69.95	66.64	67.47	96.85	95.42	94.09
(h) DAT $\lambda = 0.001$	libri 960hr mgr	99.55	99.39	98.31	**71.71**	**69.12**	**69.40**	97.05	**96.27**	**96.46**

Domain adversarial training with continual training

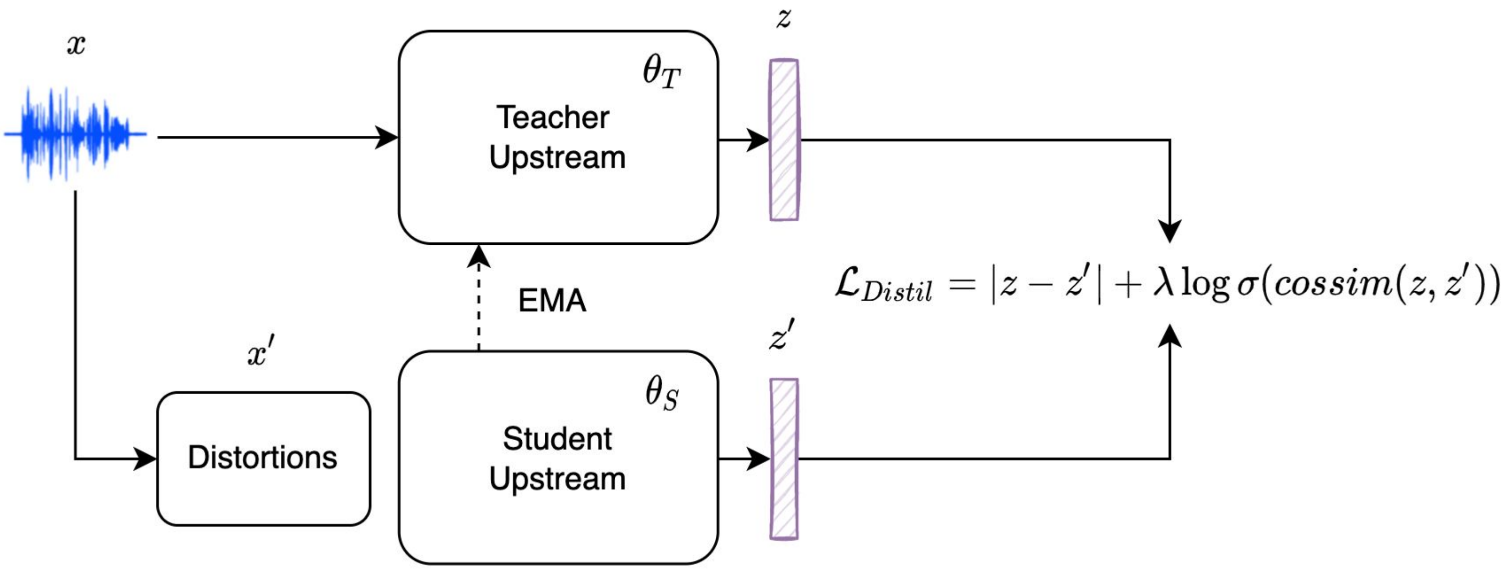
	continual	SID (Acc) clean	SID (Acc) m+g+r	SID (Acc) fsd50k	ASR (WER) clean w/o	ASR (WER) clean w/ LM	ASR (WER) m+g+r w/o	ASR (WER) m+g+r w/ LM	ASR (WER) fsd50k w/o	ASR (WER) fsd50k w/ LM	ASR (WER) CHiME3 w/o	ASR (WER) CHiME3 w/ LM
(a) baseline	-	84.97	65.51	77.61	6.72	4.88	10.16	7.94	9.62	7.57	33.4	29.26
(b) oracle	-	86.63	80.05	82.74	5.17	4.18	6.57	5.37	6.69	5.45	22.98	20.57
(c) w/o DAT	libri 100hr mgr	87.02	70.91	80.96	6.23	4.87	8.04	6.47	7.90	6.38	27.82	24.27
(d) w/o DAT	libri 960hr mgr	86.40	74.46	81.47	5.92	4.84	7.19	6.00	7.15	5.87	23.83	20.81
					Cross Entropy Loss (CE)							
(e) DAT $\lambda = 0.01$	-	86.49	71.82	79.76	6.16	4.60	11.74	9.97	11.39	12.27	44.43	41.51
(f) DAT $\lambda = 0.001$	-	87.44	72.28	79.94	6.29	5.77	9.70	9.00	9.67	8.68	32.98	29.03
					Cross Entropy Loss (CE)							
(g) DAT $\lambda = 0.001$	libri 100hr mgr	88.70	79.59	83.57	5.75	4.82	7.30	6.21	7.21	6.15	25.60	22.73
(h) DAT $\lambda = 0.001$	libri 960hr mgr	89.08	**80.27**	**85.04**	**5.49**	4.61	**6.82**	**5.69**	**6.67**	**5.62**	**23.44**	**20.71**

Future plan - Robustness of DistilHuBERT

Performance degradation for DistilHuBERT when speech has distortions.

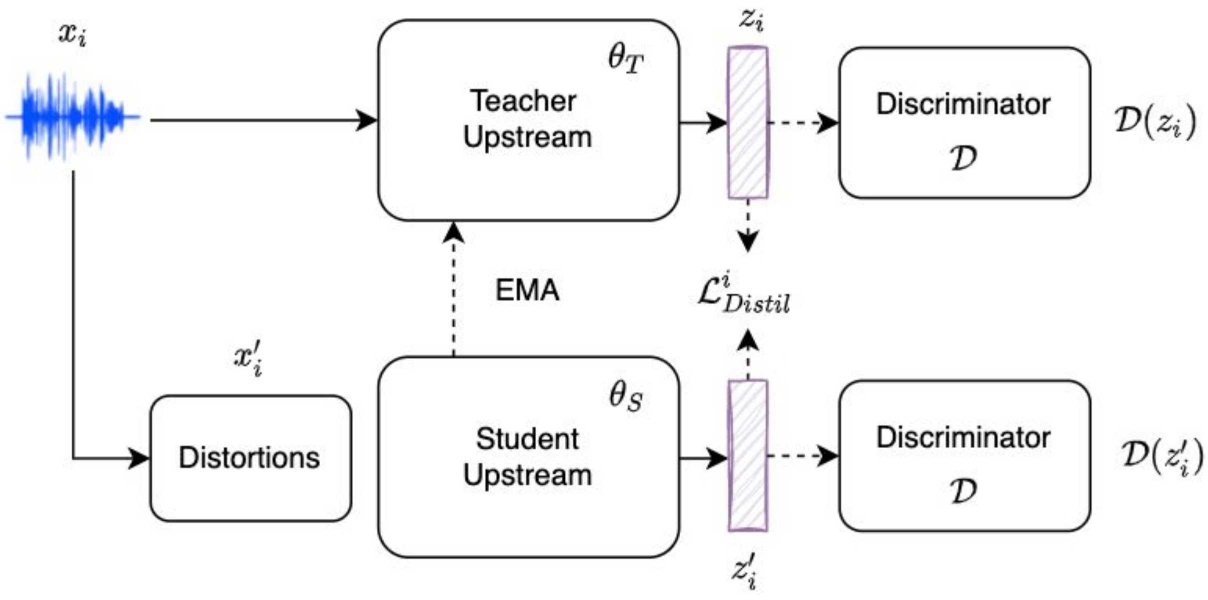
		KS	SID	IC	ER
original distilHuBERT	c	0.9516	0.6741	0.9322	0.6037
	m+g+r	0.8572	0.36	0.6167	0.5051

Future plan - Robustness of DistilHuBERT



EMA: $\theta_T \leftarrow \tau \cdot \theta_T + (1 - \tau) \cdot \theta_S$

Future plan - Robustness of DistilHuBERT



overall training objective

$$\sum_{i}^{N} \mathcal{L}^i_{Distil} + \sum_{\substack{i,\ j \\ i \neq j}}^{N} \Big(\mathcal{D}(z_i) - \mathcal{D}(z'_j)\Big)$$

$$\mathcal{L}^i_{Distil} = |z_i - z'_i|_1 + \lambda \log \sigma\Big(cossim(z_i, z'_i)\Big)$$

The End
Thanks for listening.