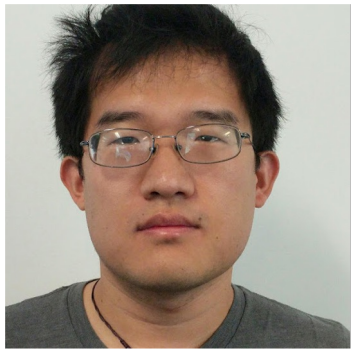
Improving Generalization Capability of Pre-trained Model

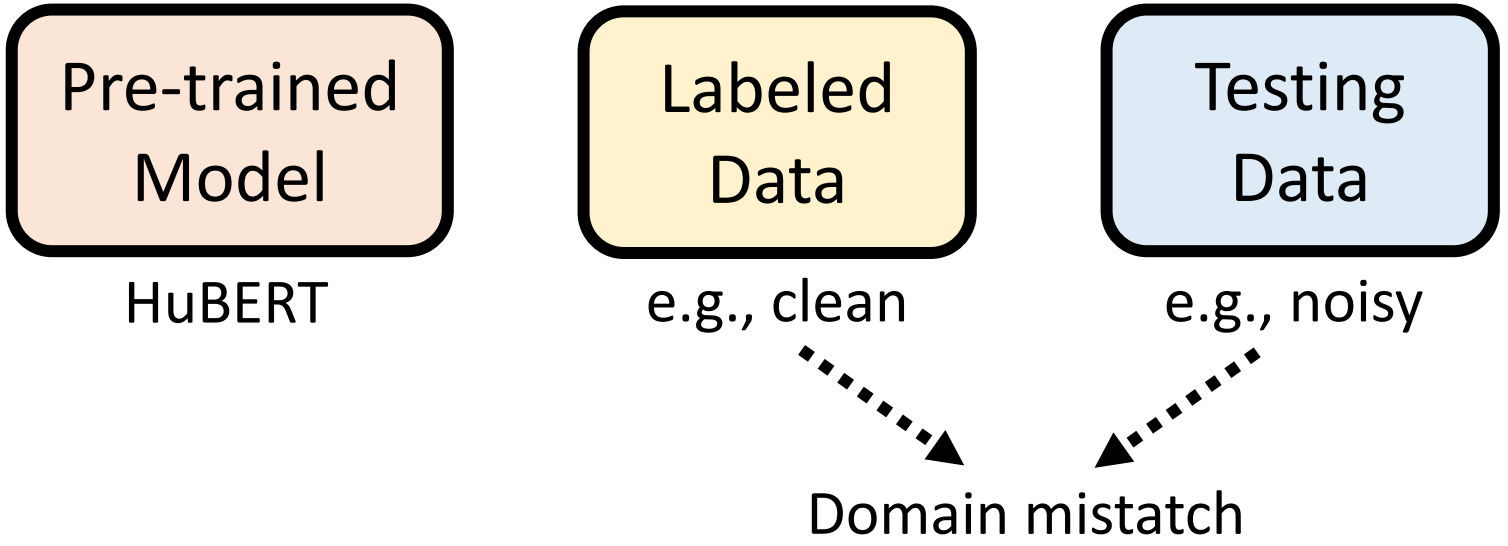
Kuan-Po Huang



Yu Zhang

Hung-yi Lee

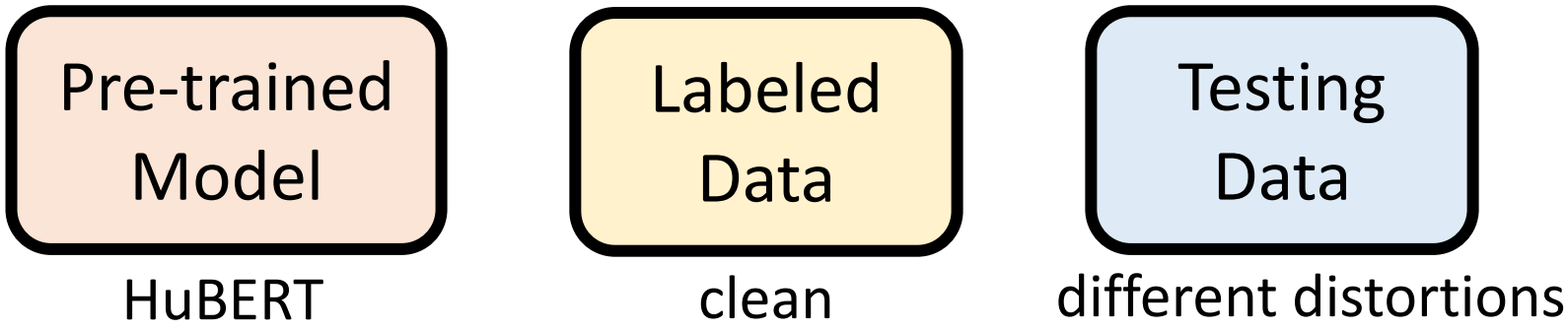
Generalization Capability of Pre-trained Model



Different domains: speech distortions, speaking styles (read vs. spontaneous), accents/dialects, languages

Focus on speech distortion at the preliminary stage

https://arxiv.org/abs/2203.16104



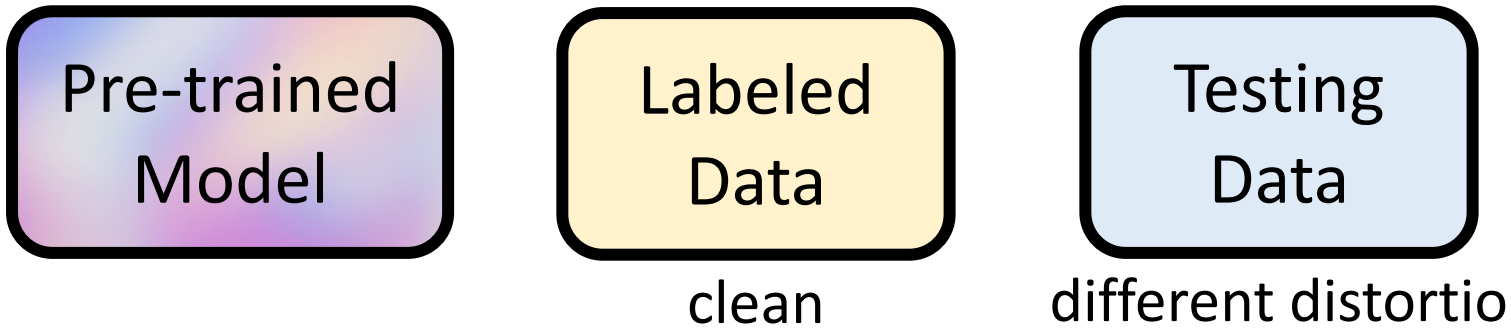
	Intent Classification ↑			Emotion Recognition ↑			Keyword Spotting ↑		
Testing Data	clean	m+g+r	fsd50k	clean	m+g+r	fsd50k	clean	m+g+r	Fsd50k
HuBERT	99.47	96.94	97.47	63.96	57.33	60.55	97.14	93.87	93.80

	Speaker Identification ↑			ASR (WER) ↓			
Testing Data	clean	m+g+r	fsd50k	clean	m+g+r	fsd50k	CHiME3
HuBERT	84.97	65.51	77.61	4.88	7.94	7.57	29.26

m+g+r = Musan + Gaussian + reverberation

https://arxiv.org/abs/2203.16104

Continuously train with noisy data (m+g+r)



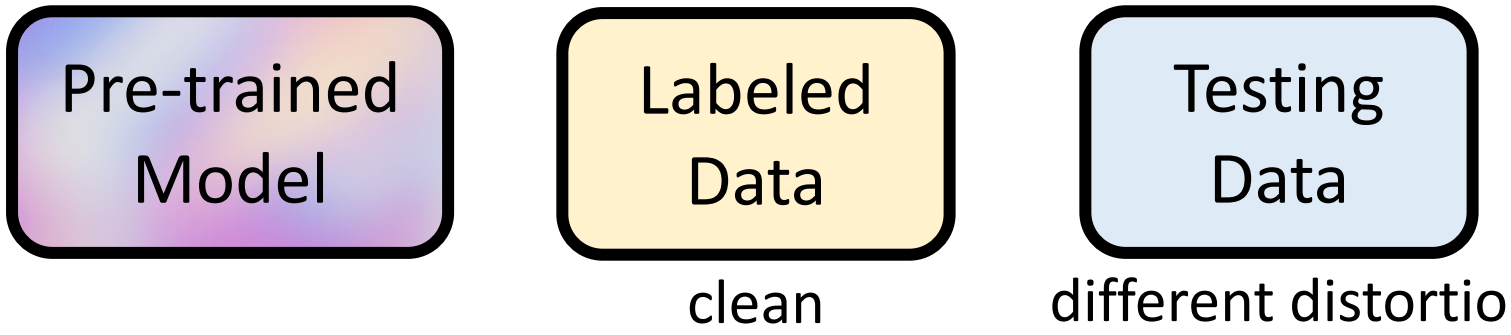
	Intent Classification ↑			Emotion Recognition ↑			Keyword Spotting ↑		
Testing Data	clean	m+g+r	fsd50k	clean	m+g+r	fsd50k	clean	m+g+r	Fsd50k
HuBERT	99.47	96.94	97.47	63.96	57.33	60.55	97.14	93.87	93.80
HuBERT + m+g+r (100hr)	99.45	98.63	97.94	64.42	62.30	60.65	96.92	94.87	93.90
HuBERT + m+g+r (960hr)	99.39	98.84	97.89	67.28	67.47	65.62	97.12	96.11	94.77

	Speaker Identification ↑			ASR (WER) ↓			
Testing Data	clean	m+g+r	fsd50k	clean	m+g+r	fsd50k	CHiME3
HuBERT	84.97	65.51	77.61	4.88	7.94	7.57	29.26
HuBERT + m+g+r (100hr)	87.02	70.91	80.96	4.87	6.47	6.38	24.27
HuBERT + m+g+r (960hr)	86.04	74.46	81.47	4.84	6.00	5.87	20.81

m+g+r = Musan + Gaussian + reverberation

https://arxiv.org/abs/2203.16104

Continuously train with noisy data (m+g+r)



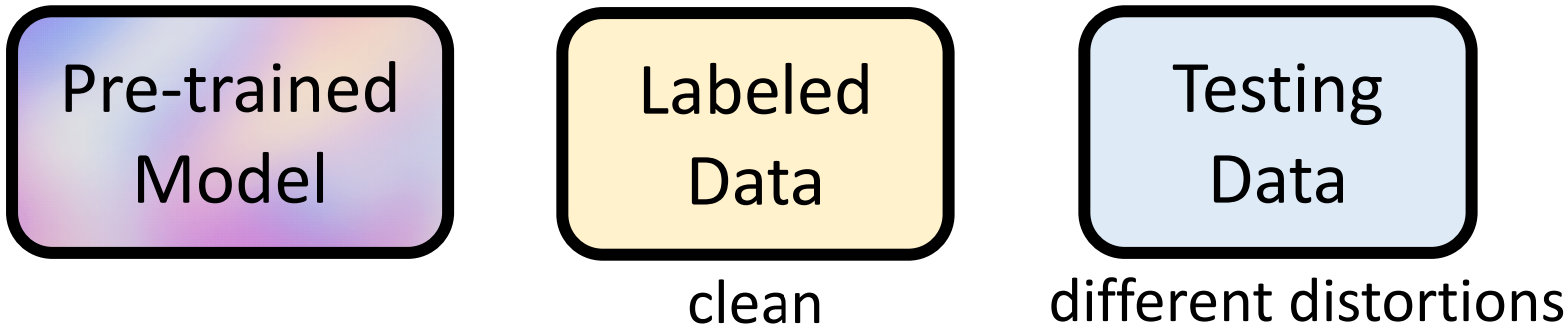
	Intent Classification ↑			Emotion Recognition ↑			Keyword Spotting ↑		
Testing Data	clean	m+g+r	fsd50k	clean	m+g+r	fsd50k	clean	m+g+r	Fsd50k
HuBERT	99.47	96.94	97.47	63.96	57.33	60.55	97.14	93.87	93.80
HuBERT + m+g+r (100hr)	99.45	98.63	97.94	64.42	62.30	60.65	96.92	94.87	93.90
HuBERT + m+g+r (960hr)	99.39	98.84	97.89	67.28	67.47	65.62	97.12	96.11	94.77

	Speaker Identification ↑			ASR (WER) ↓			
Testing Data	clean	m+g+r	fsd50k	clean	m+g+r	fsd50k	CHiME3
HuBERT	84.97	65.51	77.61	4.88	7.94	7.57	29.26
HuBERT + m+g+r (100hr)	87.02	70.91	80.96	4.87	6.47	6.38	24.27
HuBERT + m+g+r (960hr)	86.04	74.46	81.47	4.84	6.00	5.87	20.81

m+g+r = Musan + Gaussian + reverberation

https://arxiv.org/abs/2203.16104

Continuously train with noisy data (m+g+r)



Now you can use the model via S3PRL

```
model = torch.hub.load("s3prl/s3prl", "hubert_base_robust_mgr")
```

Goal: Give us a pre-trained model (compressed, visual enhanced, etc.), we make it more robust.