30 August – 3 September 2021
INTERSPEECH 2021 BRNO | CZECHIA
Speech everywhere!

AUDIO INFORMATION RESEARCH

An Empirical Study on Channel Effects for Synthetic Voice Spoofing Countermeasure Systems

You Zhang, Ge Zhu, Fei Jiang, Zhiyao Duan

you.zhang@rochester.edu

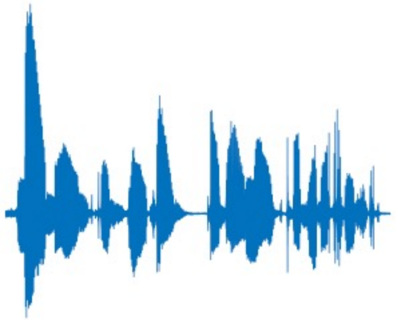
Department of Electrical and Computer Engineering,

University of Rochester, 500 Wilson Blvd, Rochester, NY 14627, USA

UNIVERSITY of ROCHESTER

Voice Biometrics

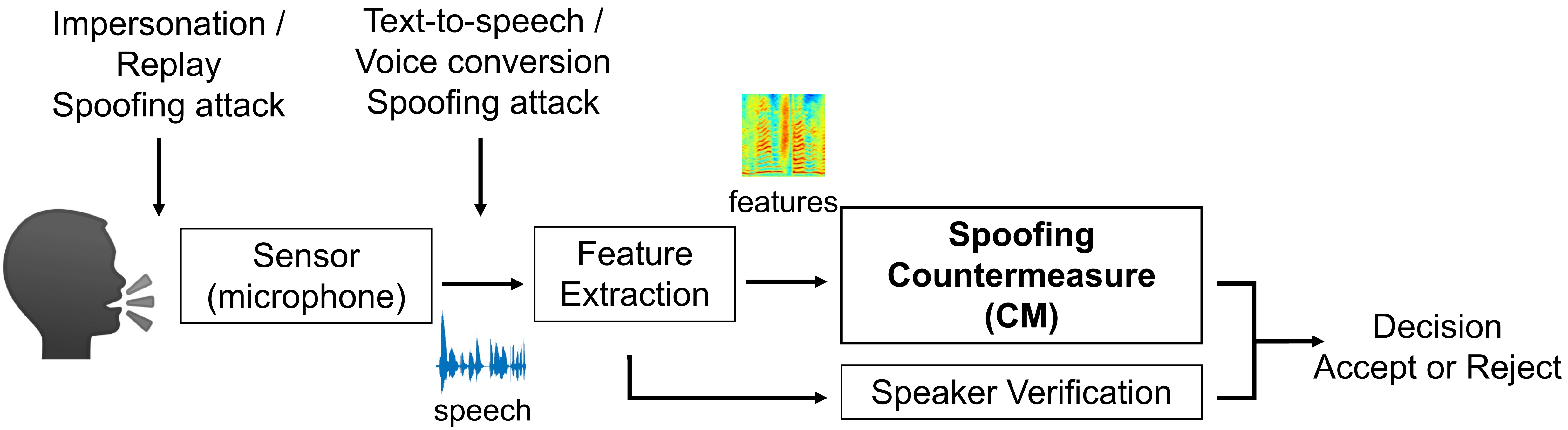
- Speaker Verification: Verify the identity of a speaker



Bona fide speech → Automatic Speaker Verification (ASV) → Decision Accept or Reject

Anti-spoofing

- Anti-spoofing / Spoofing Countermeasure: Detect spoofing attacks



ASVspoof Challenge

- Logical access (LA)
  - Text-to-speech (TTS)
  - Voice conversion (VC)
  - TTS+VC

  -- algorithm-related artifacts ★ our current focus

- Physical access (PA) -- pre-recorded, replay

  -- device-related artifacts

Outline

Background

Cross-Dataset Studies

Channel Robust Strategies

Conclusions

Background

- Spoofing countermeasure (CM) systems perform well on **single-dataset** studies.
- Several **cross-dataset** studies (trained on LA but tested on PA) have shown significant performance degradation.
- Performance degradation happens when a state-of-the-art CM system is trained and tested on **different LA datasets**.

Motivation

Reasons for Performance Degradation:

- New **attack algorithms** (**Unseen** during training)
- Other differences, such as channel variation

-- There are **limited channel effects** presented in the training set, so the CM system may fail to generalize to **unseen channel variation**.

Channel Effects

-- Audio effects imposed onto the speech signal throughout the entire recording and transmission process

- Reverberation of recording **environments**
- Frequency responses of recording **devices**
- Compression algorithms in **telecommunication**

Outline

Background

Cross-Dataset Studies

Channel Robust Strategies

Conclusions

Datasets

- ASVspoof2019LA (Training A01-A06, Evaluation A07-A19)
- ASVspoof2015 (Training S01-S05, Evaluation S01-S10)
- VCC2020 (Voice Conversion Challenge 2020): Evaluation, bona fide: training data, spoofing attacks: submitted VC systems by teams

Cross-dataset Performance

Table 1: *EER performance across different evaluation datasets (ASVspoof2019LA-eval, ASVspoof2015-eval, VCC2020). All of the three CM systems are trained on the training set of ASVspoof2019LA and validated on its development set.*

EER (%) Evaluation Datasets	CM Systems LCNN [9]	ResNet [10]	ResNet-OC [15]
2019LA-eval	3.25	5.23	2.29
2015-eval	24.55	37.11	26.30
VCC2020	33.78	36.09	41.66

- EER degradation across datasets for all three CM systems

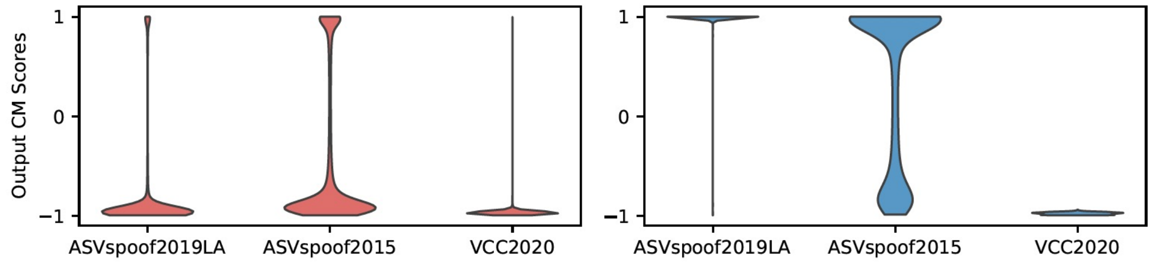


Figure 1: *Score distributions of ResNet-OC method on spoofing attacks (left) and bona fide (right) of cross-dataset evaluation.*

- The main cause is some differences in bona fide speech, among which, **channel** variation is worth checking.

Channel Mismatch

- The **average magnitude spectrum** across all bona fide utterances of each dataset is different.
- We hypothesize that channel mismatch is an important reason for the EER degradation.

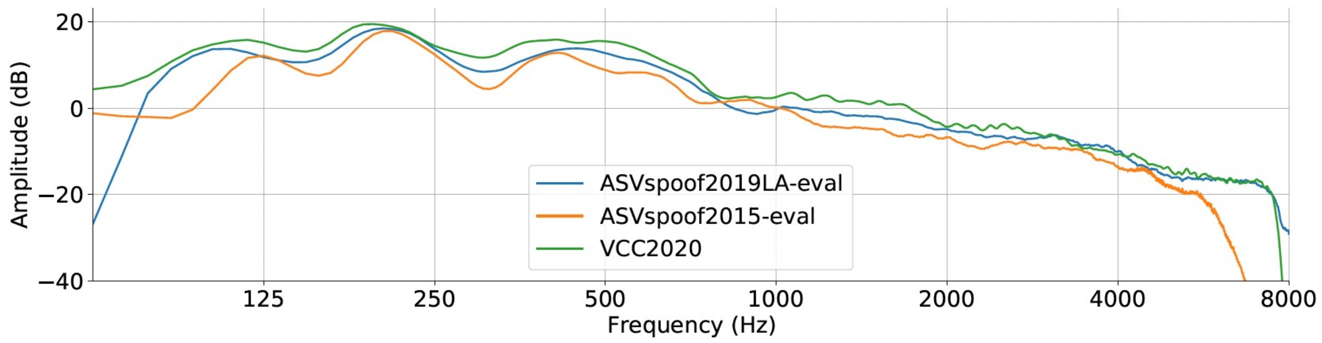


Figure 2: *Average magnitude spectra of bona fide utterances across different datasets.*

Augmentation

- ASVspoof2019LA-Sim:
  Augment ASVspoof2019LA with 12 **channel effects** by simulation

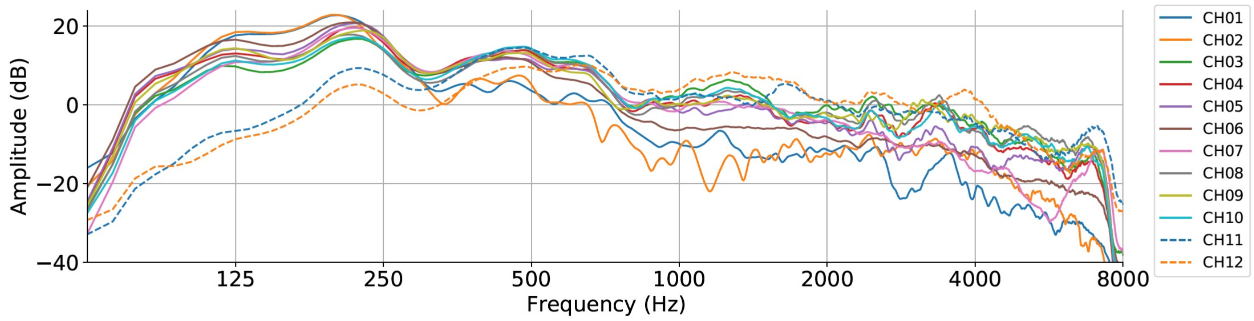


Figure 3: *Average magnitude spectra of channel-shifted bona fide utterances in the evaluation set of ASVspoof2019LA-Sim using different channel IRs.*

Results on channel-augmented Data:

- Performance degrades with channel variation, hence verifying our hypothesis.

Table 2: *EER performance on ASVspoof2019LA-Sim-eval. Average and standard deviation EERs are calculated across the 12 simulated channels. All of the three CM systems are trained on ASVspoof2019LA-train.*

EER (%) Statistics	CM Systems LCNN [9]	ResNet [10]	ResNet-OC [15]
Avg. (CH01-CH12)	27.75	48.78	40.46
Std. (CH01-CH12)	7.44	18.80	11.22

Outline

Background

Cross-Dataset Studies

Channel Robust Strategies

Conclusions

Proposed strategies

- Augmentation (AUG):
  Train with channel-augmented data (ASVspoof2019LA and 10 effects of ASVspoof2019LA-Sim)
- Multi-Task Augmentation (MT-AUG):
  Add a channel classifier
- Adversarial Augmentation (ADV-AUG):
  Insert a gradient reversal layer (GRL)

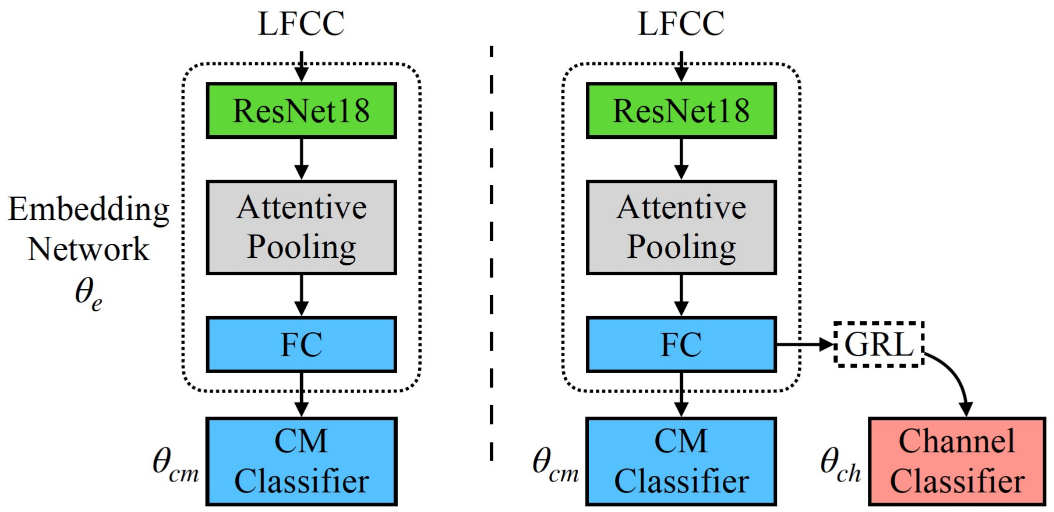


Figure 4: *Model structure of the proposed channel robust strategies. Left: Vanilla model and AUG. Right: MT-AUG (w/o GRL) and ADV-AUG (w/ GRL).*

Training objectives

- Vanilla & AUG:

$$(\hat{\theta}_e, \hat{\theta}_{cm}) = \underset{\theta_e, \theta_{cm}}{\arg\min} \mathcal{L}_{cm}(\theta_e, \theta_{cm})$$

- MT-AUG:

$$(\hat{\theta}_e, \hat{\theta}_{cm}, \hat{\theta}_{ch}) = \underset{\theta_e, \theta_{cm}, \theta_{ch}}{\arg\min} \mathcal{L}_{cm}(\theta_e, \theta_{cm}) + \lambda \mathcal{L}_{ch}(\theta_e, \theta_{ch})$$

- ADV-AUG:

$$(\hat{\theta}_e, \hat{\theta}_{cm}) = \underset{\theta_e, \theta_{cm}}{\arg\min} \mathcal{L}_{cm}(\theta_e, \theta_{cm}) - \lambda \mathcal{L}_{ch}(\theta_e, \hat{\theta}_{ch})$$

$$(\hat{\theta}_{ch}) = \underset{\theta_{ch}}{\arg\min} \mathcal{L}_{ch}(\hat{\theta}_e, \theta_{ch})$$

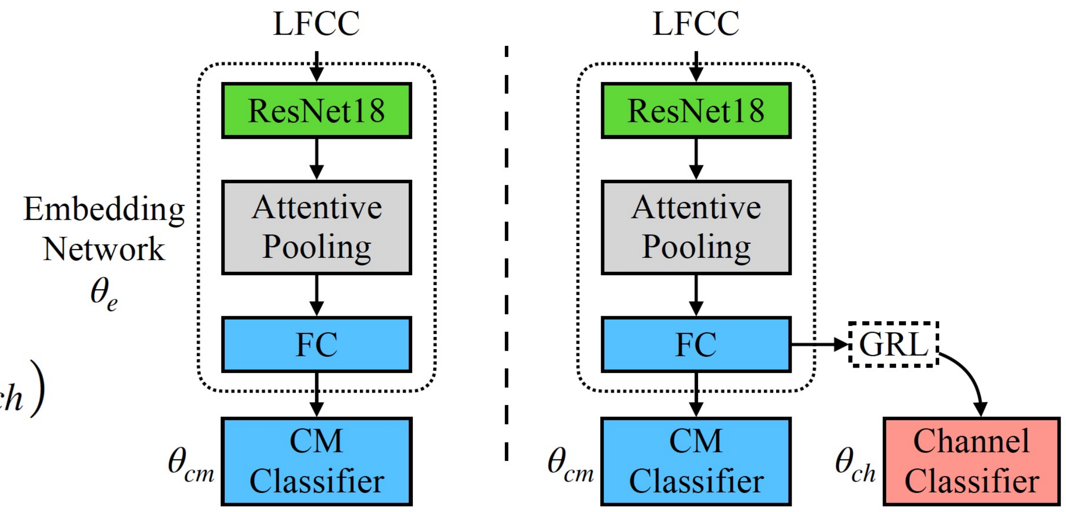


Figure 4: *Model structure of the proposed channel robust strategies. Left: Vanilla model and AUG. Right: MT-AUG (w/o GRL) and ADV-AUG (w/ GRL).*

In-domain test

Table 3: *EER performance comparison of the proposed strategies and the vanilla model on ASVspoof2019LA-Sim-eval. The proposed strategies are trained on the augmented training set.*

EER (%)	Methods			
	Vanilla	AUG	MT-AUG	ADV-AUG
Avg. (CH01-10)	38.14	4.43	4.29	3.92
Std. (CH01-10)	10.83	0.75	0.46	0.43
CH 11	54.98	3.58	4.59	3.78
CH 12	49.17	4.41	7.08	6.28

- Test on channel-augmented data
- The strategies make the CM system less sensitive to channel variation

Out-of-Domain Test

- Our proposed channel-robust strategies show significant improvement on both out-of-domain datasets, ASVspoof2015-eval and VCC2020
- Verified our hypothesis of channel mismatch among these datasets
- Verified the effectiveness of the proposed strategies

Table 4: *EER comparison of the proposed strategies and the vanilla model on cross-dataset evaluation.*

EER(%) Evaluation Datasets	Methods Vanilla	AUG	MT-AUG	ADV-AUG
2019LA-eval	2.29	2.92	3.41	3.23
2015-eval	26.30	16.25	22.10	14.38
VCC2020	41.66	30.51	28.85	27.07

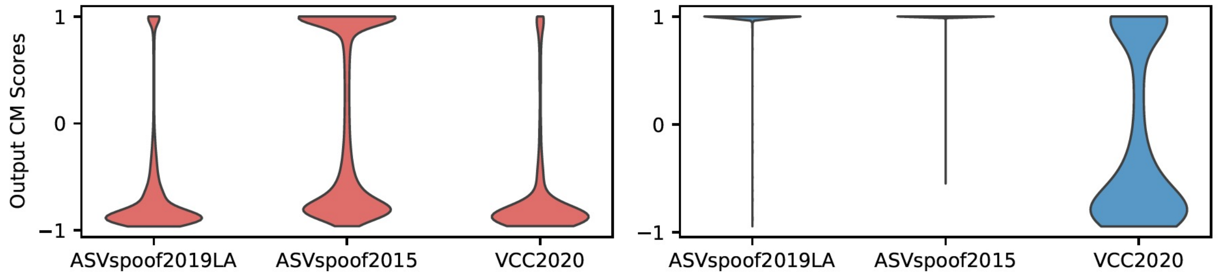


Figure 6: *Score distributions of ADV-AUG strategy on spoofing attacks (left) and bona fide (right) of cross-dataset evaluation.*

Outline

Background

Cross-Dataset Studies

Channel Robust Strategies

Conclusions

Conclusions

- The channel mismatch between training and evaluation is indeed an important reason for the performance degradation of CM systems.

- Our proposed several strategies (data augmentation, multi-task learning, adversarial learning) improve the robustness of CM systems to channel variation.

AUDIO INFORMATION RESEARCH

- Our code will be available at https://github.com/yzyouzhang/Empirical-Channel-CM

Code and data

- Feel free to check out our follow-up paper in ASVspoof 2021 Workshop: https://arxiv.org/pdf/2107.12018.pdf

AUDIO INFORMATION RESEARCH

Thank you !

Q & A

you.zhang@rochester.edu