A satellite event of INTERSPEECH 2024

ASVspoof

Automatic Speaker Verification and Spoofing Countermeasures Challenge

ASVspoof Workshop 2024

Malacopula: adversarial automatic speaker verification attacks using a neural-based generalised Hammerstein model

Massimiliano Todisco, Michele Panariello, Xin Wang,
Héctor Delgado, Kong Aik Lee, Nicholas Evans

31 August 2024 - Kos Island, Greece

Intro

Advancements and Challenges in Automatic Speaker Verification (ASV) Systems

- Recent ASV systems demonstrate greater robustness to spoofing attacks
- ASV systems remain vulnerable to adversarial attacks
- Adversary can exploit ASV weaknesses, leading to false verifications

Objectives

- Present **Malacopula**[1], an adversarial attack against ASV to make spoofs more effective
- Based on the generalised Hammerstein model
- Based on convolutive noise that modifies the signal in a non-linear fashion

Security Concerns

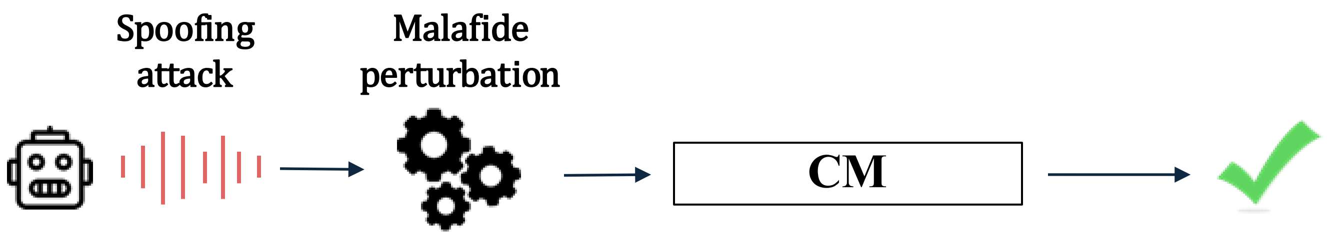
- Malacopula increases ASV vulnerabilities
- Requires minimal computational resources
- Adds noises that can be mistaken for real, common sounds
- Highlights the need for stronger defences in real-world scenarios

[1] *Mala copula* is Latin for "bad connection" or "bad union. It signifies an undesirable or improper association between elements.

Adversarial attacks on ASV

Previous works

- **Fixed noise pattern** [1]: generation of utterance-dependent adversarial noise
- **Utterance length dependent** [2]: concatenated adversarial noise pattern
- **ASV-dependent** [3]: operates within a white-box threat model, where the adversary has full knowledge of the ASV system
- **Malafide [4]**
  - Early work which serves as the inspiration for Malacopula
  - Malafide leverages a convolutional filter applied directly to the raw waveform, optimised to manipulate a countermeasure (CM) system

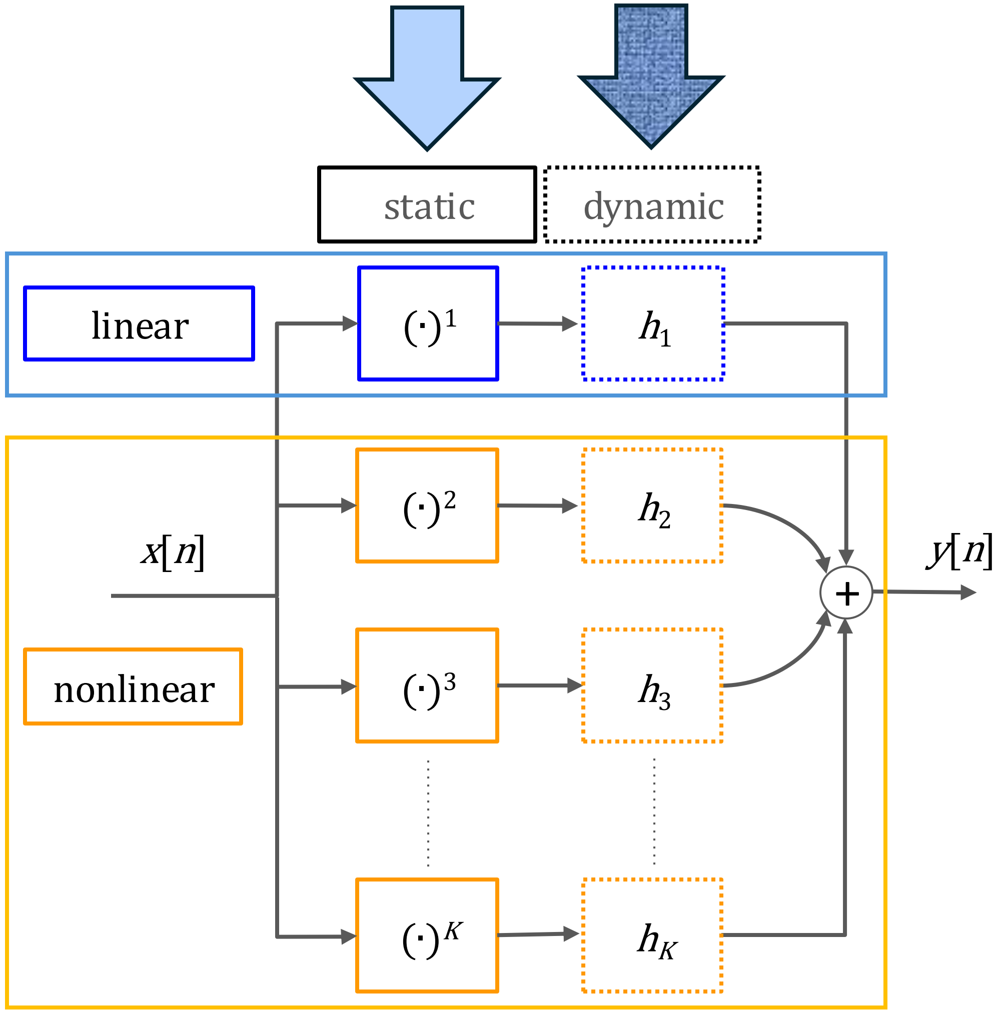


[1] X. Zhang et al., “Waveform level adversarial example generation for joint attacks against both automatic speaker verification and spoofing countermeasures,” in Engineering Applications of AI, 2022.
[2] Yi Xie, CongShi, Zhuohang Li, Jian Liu, Yingying Chen, and Bo Yuan, “Real-time, universal, and robust adversarial attacks against speaker recognition systems,” in Proc. ICASSP 2020.
[3] Weiyi Zhang, et al, “Attack on practical speaker verification system using uni- versal adversarial perturbations,” in Proc. ICASSP 2021.
[4] Michele Panariello, Wanying Ge, Hemlata Tak, Massimiliano Todisco, and Nicholas Evans, “Malafide: a novel adversarial convolutive noise attack against deepfake and spoofing detection systems,” 2023.

The Generalised Hammerstein Model

- **Non-linear transformation**: Captures the non-linear characteristics of the input signal using K static polynomial functions $\phi_k(\cdot) = (\cdot)^k$
- **Linear Time-Invariant (LTI) filters**: Applies dynamic, K linear filters h_k of length L to process the transformed signal

$$y[n] = \sum_{k=1}^{K} \sum_{i=0}^{L-1} h_k[i] \cdot \phi_k(x[n-i])$$



Applications [1-3]

- Widely used in audio processing, acoustics, and modelling of complex systems
- Offers flexibility and computational efficiency for handling non-linear distortions in various domains

[1] Simon Grimm and Jurgen Freudenberger, "Hybrid Volterra and Hammerstein modelling of nonlinear acoustic systems," in Fortschritte der Akustik: DAGA 2016.
[2] Giovanni L. Sicuranza and Alberto Carini, "On the accuracy of generalized Hammerstein models for nonlinear active noise control," in Proc. 2006 IEEE IMT Conference, 2006.
[3] Hemlata Tak et al, "RawBoost: A raw data boosting and augmentation method applied to automatic speaker verification anti-spoofing," in Proc. ICASSP 2022.

Malacopula: Overview

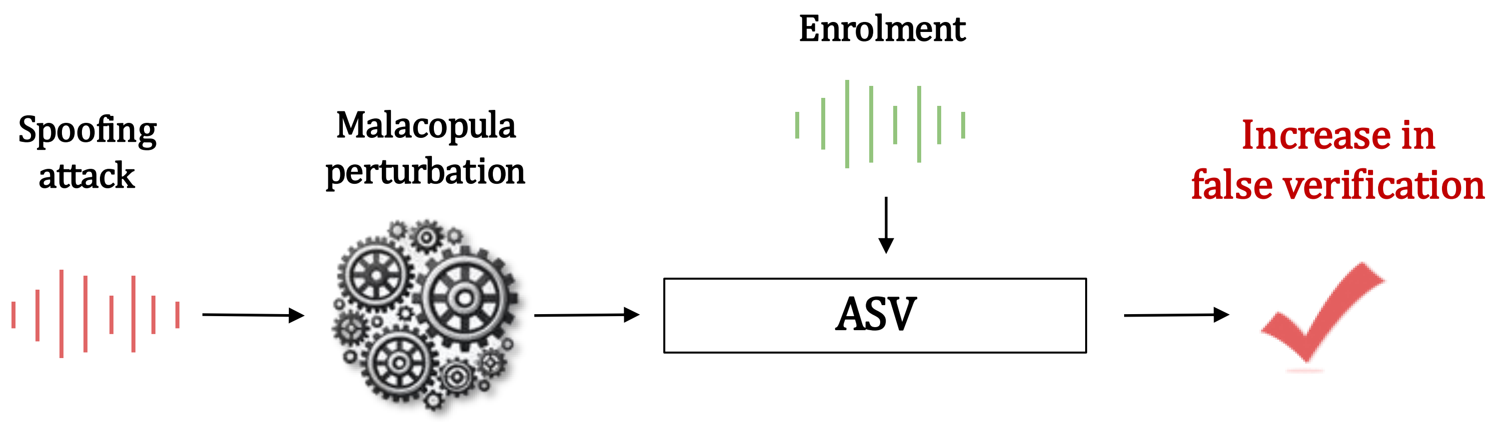
Automatic Speaker Verification and Spoofing Countermeasures Challenge

Purpose

- Develop adversarial attacks to increase false ASV verifications by enhancing spoofing attacks

Key Innovation

- **Neural-based** generalised Hammerstein model with learnable parameters
- **Speaker- and attack-specific**
- **Transferable** across different ASV systems and utterances
- **Invariant** to utterance duration or content
- **Lightweight** optimisation, making it efficient, easy to deploy and use
- Acts as a **post-processing filter** which applies adversarial perturbations to spoofed speech

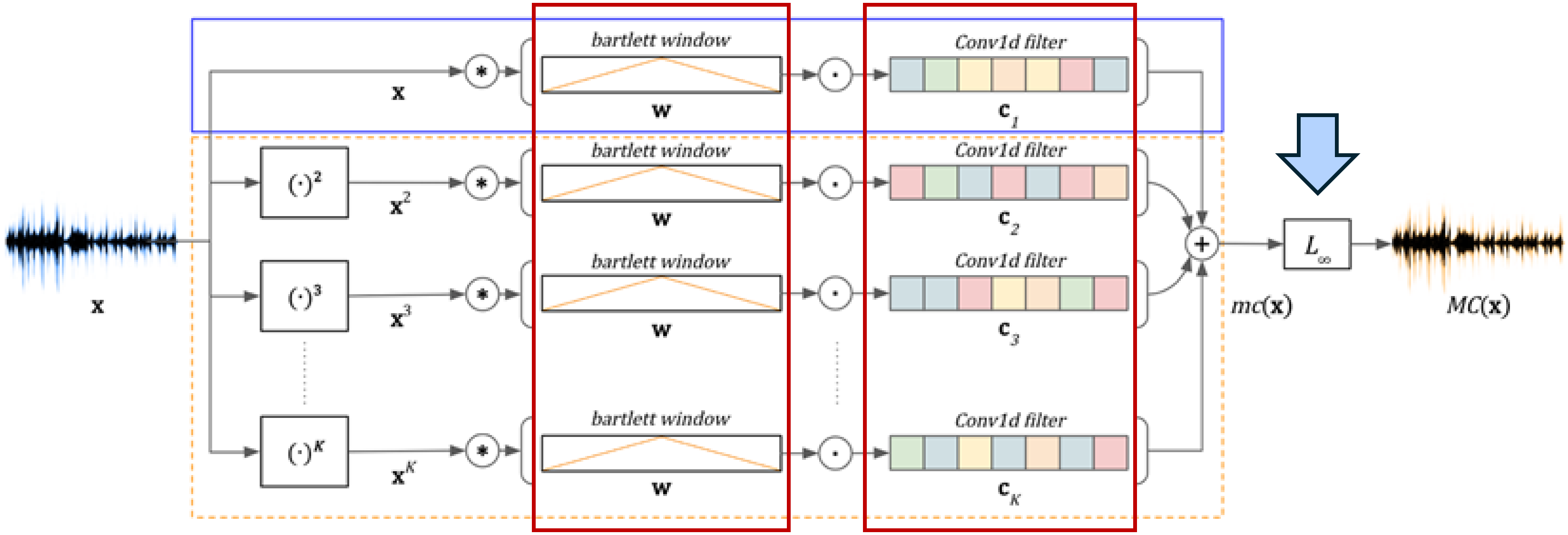


Malacopula Filter Architecture

Automatic Speaker Verification and Spoofing Countermeasures Challenge

Filter Design

- Multiple parallel branches with different levels of non-linearity
- Bartlett window to reduce spectral leakage while balancing frequency resolution and dynamic range
- A normalisation layer using the L_∞ norm is applied after summation to prevent distortion



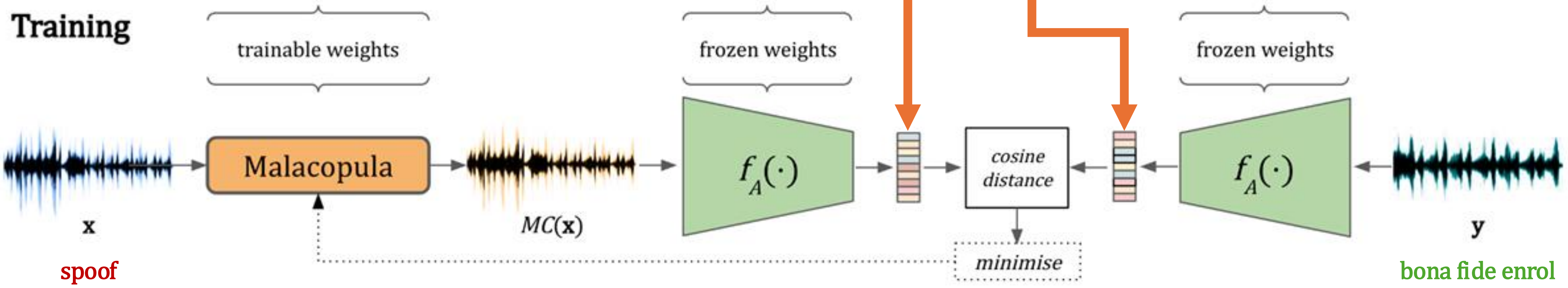
$$mc_{K,L}(\mathbf{x}) = \sum_{k=1}^{K} \left[\mathbf{x}^k * (\mathbf{w} \odot \mathbf{c}_{k,L}) \right] \longrightarrow MC(\mathbf{x}) = \frac{mc(\mathbf{x})}{|mc(\mathbf{x})|_\infty}$$

Adversarial Optimisation Procedure (1/2)

Training Process

- **Objective**: minimise the *cosine distance* between embeddings of perturbed spoofed utterances and bona fide enrolment utterance
- $f_A(\cdot)$ is the speaker embedding extractor
- Each Malacopula filter is trained independently for a specific speaker s and a spoofing algorithm a

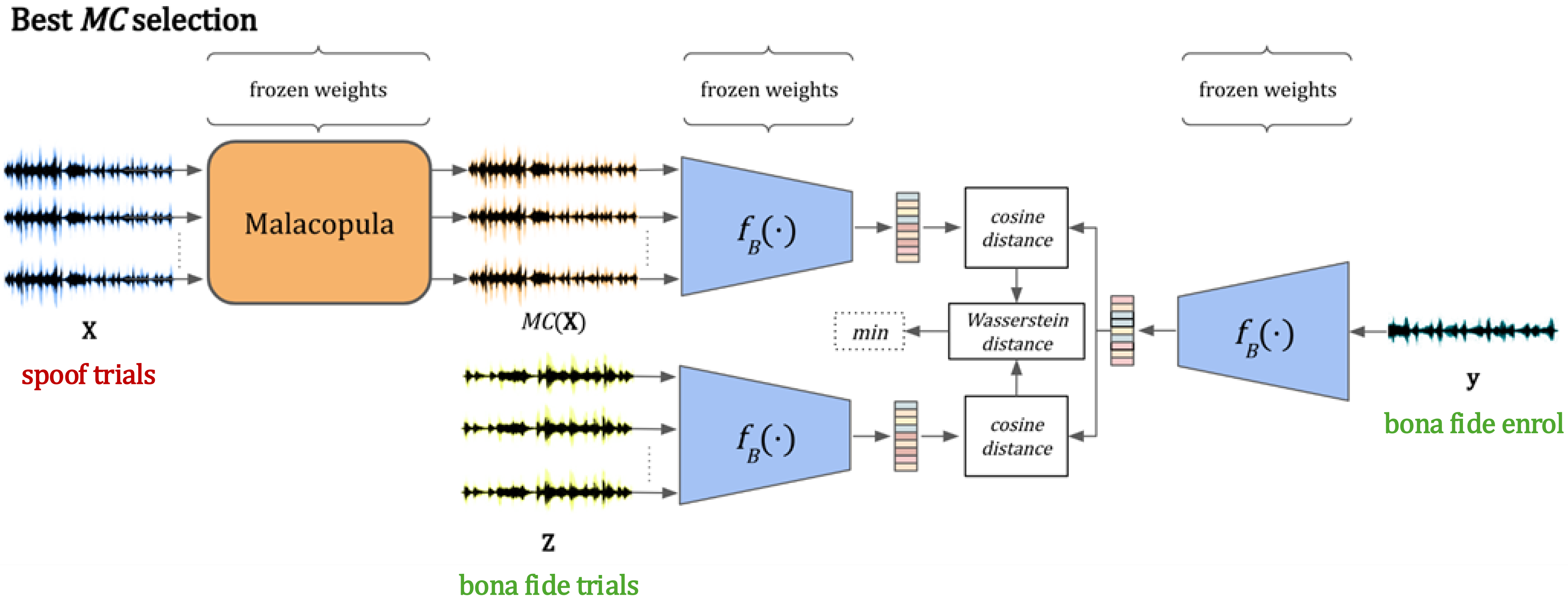
$$\min_{\mathbf{c}_{K,L}^{(s,a)}} \left[1 - CS\left(f_A\left(MC_{K,L}^{(s,a)}(\mathbf{x})\right), f_A\left(\mathbf{y}\right)\right)\right]$$



Adversarial Optimisation Procedure (2/2)

Validation Process

- **Cross-System Generalisation**: to ensure robustness across different ASV systems, the best filter is selected using a second speaker embedding extractor $f_B(\cdot)$
- **Selection Criterion**: based on the minimum Wasserstein distance to capture full distribution characteristics between spoofed and bona fide score distributions



Experimental Setup

ASV systems

- CAM++ [1] (**Training**), ECAPA [2] (**Validation**) and ERes2Net [3] (**Testing**)

Dataset and Attack Scenario

- ASVspoof 2019 Logical Access [4]
- Malacopula filters were trained offline using evaluation partition data → attacks from $\mathrm{A07}$ to $\mathrm{A19}$
- Filters were optimised for 48 speakers and 13 spoofing algorithms

Optimisation Details

- Adam optimiser used with 60 epochs and a batch size of 12
- Filters explored with two different lengths: $L = \{257, 1025\}$
- Filter depths tested: $K = \{1, 3, 5\}$

Evaluation

- Results computed using the standard SASV [5] protocol, expressed as spf-EERs computed using target and spoofed utterances

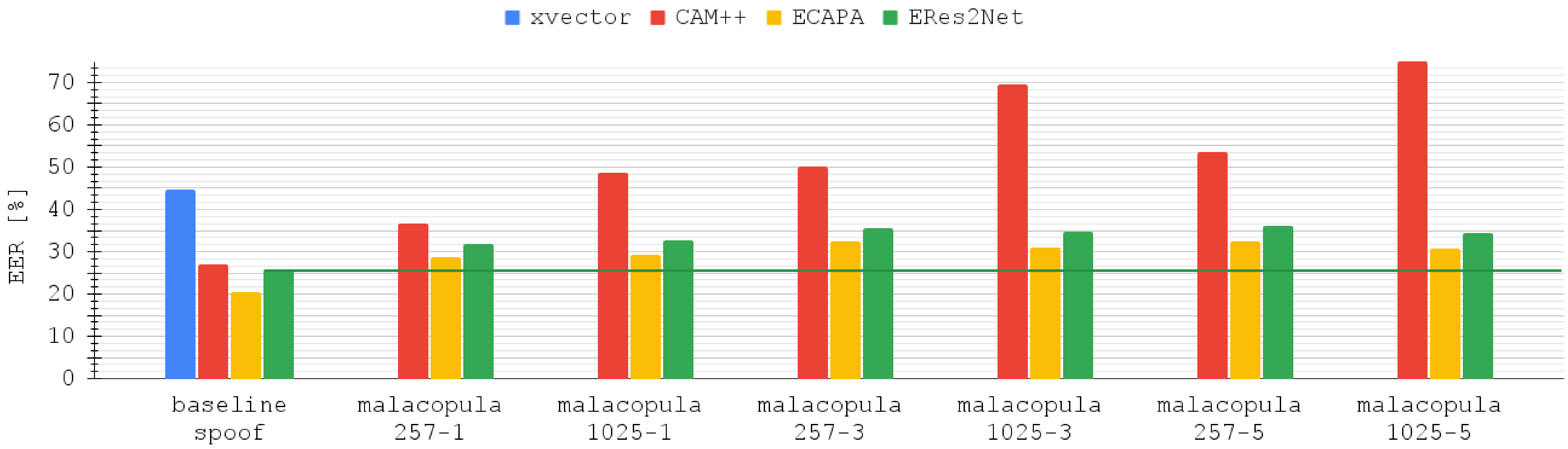
[1] Haibo Wang, Siqi Zheng, Yafeng Chen, Luyao Cheng, and Qian Chen, "CAM++: A fast and efficient network for speaker verification using context-aware masking," in Proc. INTERSPEECH 2023, 2023.
[2] Brecht Desplanques, Jenthe Thienpondt, and Kris Demuynck, "ECAPA-TDNN: Emphasized channel attention, propagation and aggregation in TDNN based speaker verification," in Proc. INTERSPEECH 2020.
[3] Yafeng Chen et al., "ERes2NetV2: Boosting short- duration speaker verification performance with computational efficiency," arXiv preprint arXiv:2406.02167, 2024.
[4] Xin Wang et al., "ASVspoof 2019: A large-scale public database of synthesized, converted and replayed speech," Computer Speech & Language, vol. 64, pp. 101114, 2020.
[5] J.-w. Jung, H. Tak, H.-j. Shim, H.-S. Heo, B.-J. Lee, S.- W. Chung, H.-J. Yu, N. Evans, and T. Kinnunen, "SASV 2022: The first spoofing-aware speaker verification challenge," in INTERSPEECH 2022, 2022

Experimental Results: ASV System Vulnerabilities

ASVspoof Automatic Speaker Verification and Spoofing Countermeasures Challenge

Increased Vulnerabilities

- Malacopula significantly increases ASV system vulnerabilities across all tested systems
- Greatest impact observed with CAM++
- Malacopula also increases vulnerabilities of ECAPA and ERes2Net systems
- spf-EER results indicate effective attack generalisation across different ASV architectures

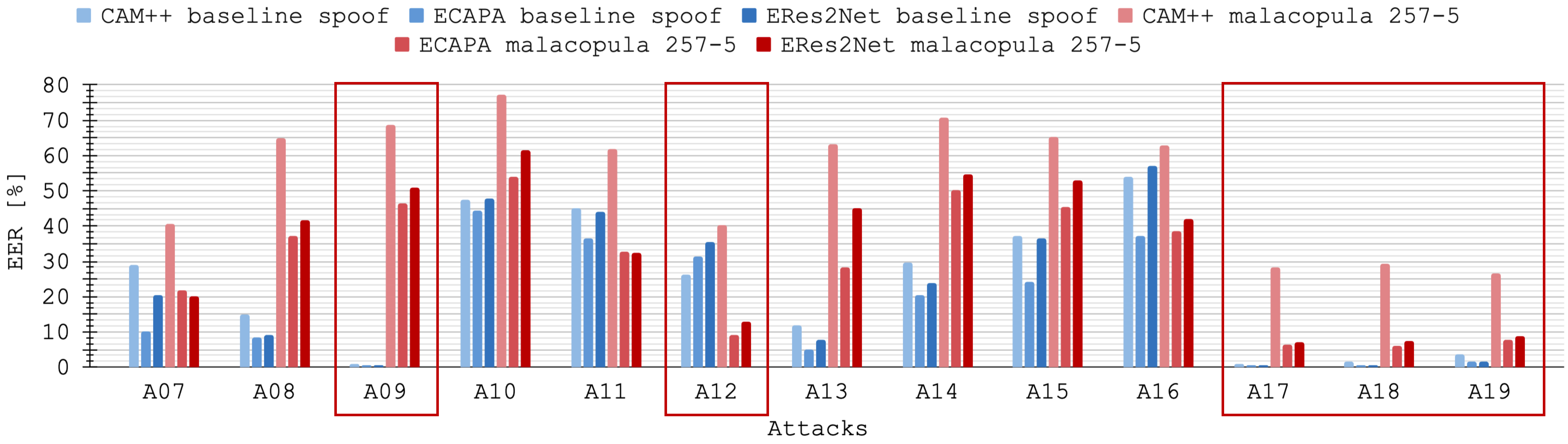


Experimental Results: Attack Specific Performance

ASVspoof
Automatic Speaker Verification and Spoofing Countermeasures Challenge

Attacks Breakdown

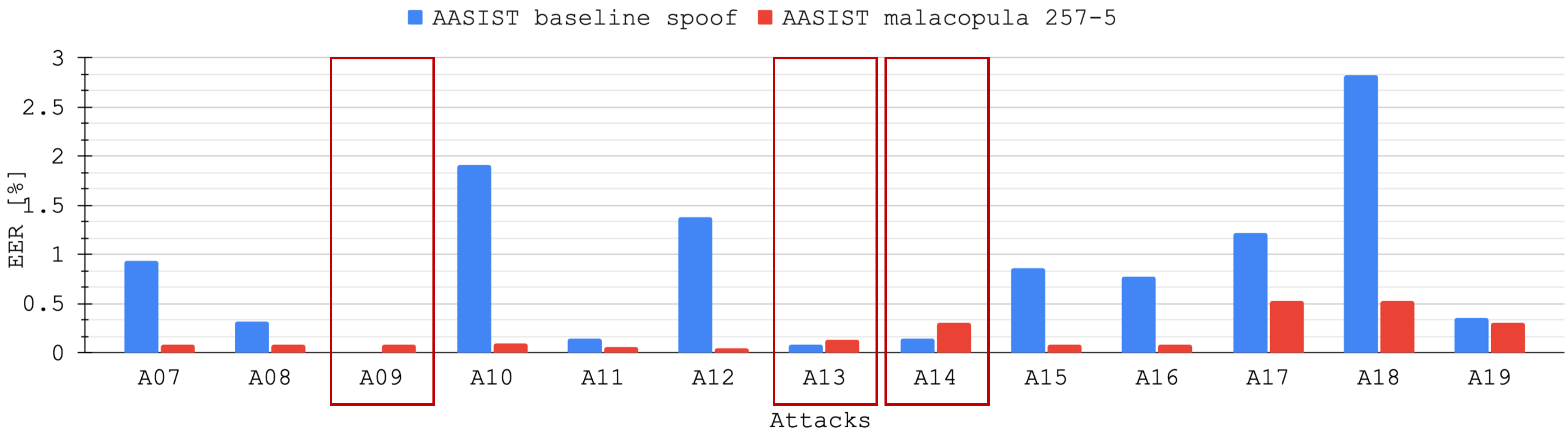
- Impact of Malacopula ($L = 257, K = 5$) varies across different spoofing attacks
- High impact for certain attacks (e.g., $A09$) and lower impact for others (e.g., $A12$)
- High impact for voice conversion attacks $A17$, $A18$ and $A19$



Impact upon Anti-spoofing Detection

AASIST [1] Performance

- Global improvement in detection performance for most attacks, with a few exceptions (e.g., A09, A13 and A14), where spf-EERs slightly increase
- Further research is required to validate detection performance in real-world, unconstrained scenarios, where conditions differ from controlled environments

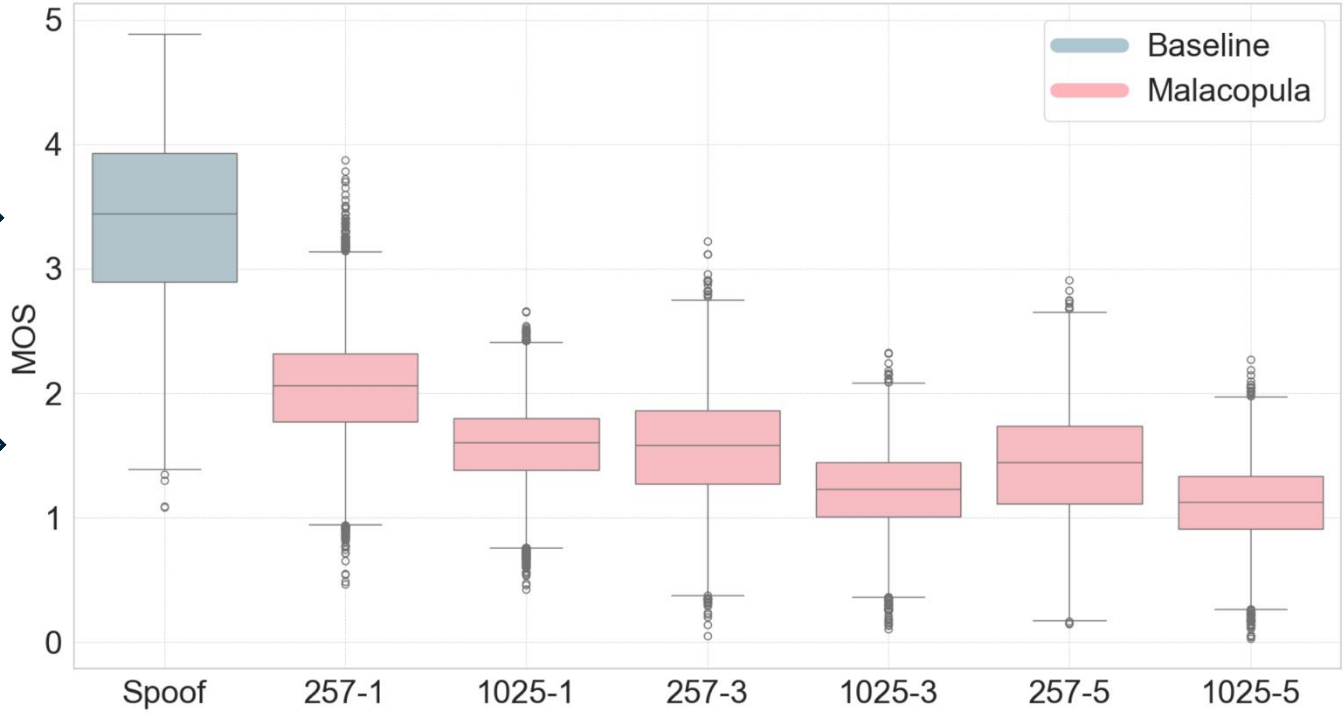


[1] Jee-weon Jung, et al, "AASIST: Audio anti-spoofing using integrated spectro-temporal graph attention networks," in Proc. ICASSP 2022.

Impact on Speech Quality

Speech Quality MOS Performance [1]

- Reduction in speech quality, as reflected by lower Mean Opinion Scores (MOS)
- Smaller filter configurations cause less degradation, while larger lead to greater degradation
- Though Malacopula adds noise, it can be mistaken for real, common sounds
- Detection challenging despite quality degradation



[1] Erica Cooper, Wen-Chin Huang, Tomoki Toda, and Junichi Yamagishi, "Generalization ability of MOS prediction networks," in Proc. ICASSP 2022.

Malacopula Audio Examples

Speaker ID L0007		
	Bona fide enrolment	
A09	A13	A17
Malacopula (L =257, K=5)	Malacopula (L =257, K=5)	Malacopula (L =257, K=5)

Conclusions

Increased ASV Vulnerabilities

- Malacopula significantly increases ASV system vulnerabilities by enhancing spoofing attacks through non-linear perturbations

Adaptability and Efficiency

- The lightweight model adapts efficiently across different ASV systems, attack scenarios and utterance lengths, making it both versatile and effective in various settings

Anti-spoofing Detection

- Detection systems like AASIST can still effectively identify Malacopula attacks **BUT** under controlled conditions

Speech Quality

- While Malacopula reduces speech quality, the added noise can be mistaken for real sounds, making detection challenging

Security Risks

- Further investigation is needed to assess ASV robustness in real-world, unconstrained environments

Malacopula Project on GitHub

Access Malacopula Code and Resources

- GitHub Link:
  - https://github.com/eurecom-asp/malacopula

What's Available

- Complete implementation of the Malacopula model
- Audio examples and resources to reproduce results
- Detailed instructions for running experiments and applying the model

Strong defences start with strong attacks

Thank you