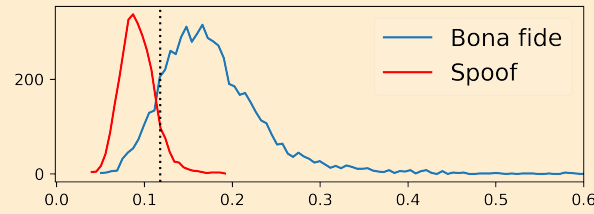


Distribution of feature F85 (mean unvoiced length) on attack A10 and bona fide data from ASVspoof5



🎯 Goal · Detect synthetically generated audio using hand-crafted acoustic features.

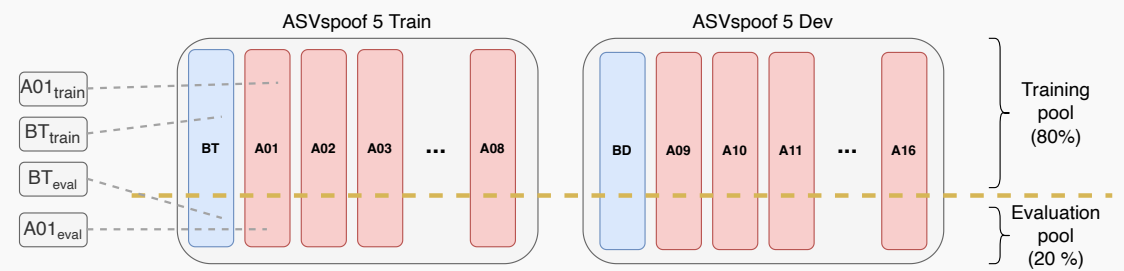
💡 Finding 1 · Even single features perform surprisingly well.

💡 Finding 2 · Generalisation across attacks remains challenging; this suggests potential shortcuts in the data.

openSMILE eGeMAPS features

- Minimalistic set of 88 acoustic parameters
- Used for voice analysis and affect recognition
- Includes energy, spectral, temporal features
- Summarized with mean, std, percentiles
- Examples of predictive features:
 - F19**: standard deviation of the falling slope of the moving average of the loudness
 - F21**: standard deviation-normalized moving average of the spectral flux
 - F66**: mean of the moving average of the spectral flux for non-zero values
 - F85**: mean length of all unvoiced segments within the audio file
 - F86**: standard deviation of the length of all unvoiced segments.

Dataset: ASVspoof 5

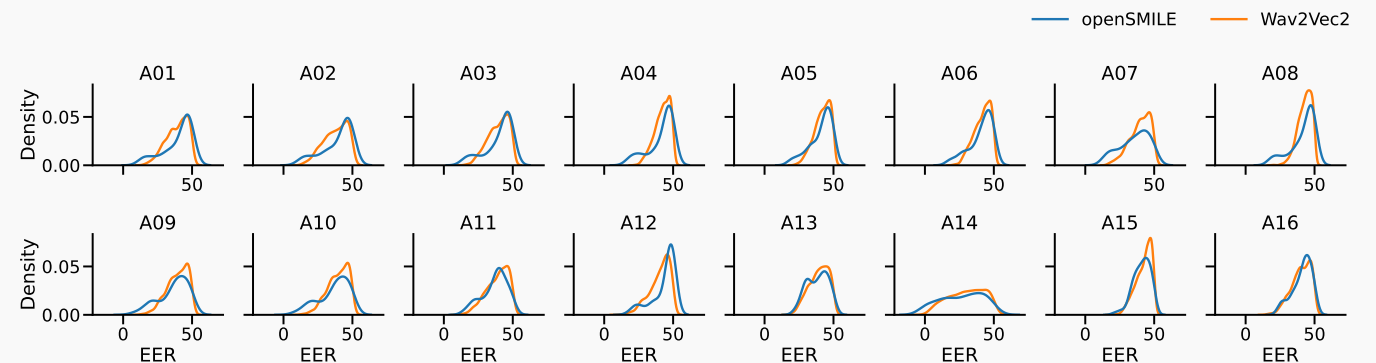


Attack	System	Attack	System
A01	GlowTTS	A09	ToucanTTS
A02	Variant of A01	A10	A09+HifiGANv2
A03	Variant of A01	A11	Tacotron2
A04	GradTTS	A12	In-house unit-select
A05	Variant of A04	A13	StarGANv2-VC
A06	Variant of A04	A14	YourTTS
A07	FastPitch	A15	VAE-GAN
A08	VITS	A16	In-house ASR-based

Main results

- Linear classification model
- Feature selection:
 - best: single best feature for each attack
 - all: 88 for openSMILE, 768 for Wav2Vec2
- Train on BT_{train} , BD_{train} , Ai_{train}
- Evaluate on ...
 - BT_{eval} , BD_{eval} , Ai_{eval} (in domain)
 - BT_{eval} , BD_{eval} , Aj_{eval} , $j \neq i$ (out of domain)
- Metric: Equal error rate (EER) ↓

Performance distribution of scalar features: openSMILE vs. Wav2Vec2



Results of the best scalar feature identified on each attack

Attack	In domain			Out of domain		
	openSMILE best	W2V2 all	all	openSMILE best	W2V2 all	all
A01	12.8	1.6	0.0	45.1	41.9	5.5
A02	12.7	0.6	0.0	45.1	40.2	4.8
A03	15.3	1.7	0.0	41.4	38.9	4.7
A04	16.8	1.4	0.3	41.3	33.9	14.3
A05	17.5	1.0	0.3	28.4	26.7	15.6
A06	16.8	1.3	0.2	28.5	27.5	13.5
A07	13.7	0.3	0.0	45.0	37.2	13.1
A08	17.7	5.0	0.0	50.3	41.6	9.3
A09	8.0	2.6	0.0	35.2	34.7	13.1
A10	7.3	2.1	0.2	35.4	35.2	11.1
A11	15.3	0.7	0.0	46.3	30.7	14.8
A12	18.2	8.5	0.2	46.0	50.1	25.3
A13	22.1	2.5	0.5	32.5	33.7	16.0
A14	0.8	0.0	0.0	54.1	38.5	18.8
A15	20.7	6.5	0.5	52.6	41.6	8.1
A16	23.8	3.3	0.0	59.2	33.8	9.1

Attack	Feature	ASVspoof 5 Train								ASVspoof 5 Dev							
		A01	A02	A03	A04	A05	A06	A07	A08	A09	A10	A11	A12	A13	A14	A15	A16
A01	F66	12.8	12.7	15.7	21.7	31.5	25.7	13.7	18.2	68.7	69.5	68.1	70.3	60.0	55.2	68.7	76.2
A02	F66	12.8	12.7	15.7	21.7	31.5	25.7	13.7	18.2	68.7	69.5	68.1	70.3	60.0	55.2	68.7	76.2
A03	F19	13.8	13.0	15.3	16.8	19.4	17.1	17.8	23.0	61.9	62.8	55.7	78.7	54.7	54.5	65.3	66.7
A04	F19	13.8	13.0	15.3	16.8	19.4	17.1	17.8	23.0	61.9	62.8	55.7	78.7	54.7	54.5	65.3	66.7
A05	F21	21.9	21.6	24.8	17.3	17.5	16.8	18.7	45.7	18.1	17.5	40.0	43.7	33.8	28.6	36.5	41.5
A06	F21	21.9	21.6	24.8	17.3	17.5	16.8	18.7	45.7	18.1	17.5	40.0	43.7	33.8	28.6	36.5	41.5
A07	F66	12.8	12.7	15.7	21.7	31.5	25.7	13.7	18.2	68.7	69.5	68.1	70.3	60.0	55.2	68.7	76.2
A08	F20	17.2	17.2	18.8	28.2	36.1	30.9	19.9	17.7	77.7	79.0	69.2	70.9	60.8	89.0	67.5	72.7
A09	F86	26.3	22.3	23.9	28.2	23.3	24.2	20.6	49.8	8.0	7.3	61.5	38.0	49.0	65.0	45.5	43.3
A10	F86	26.3	22.3	23.9	28.2	23.3	24.2	20.6	49.8	8.0	7.3	61.5	38.0	49.0	65.0	45.5	45.0
A11	F58	48.6	50.3	49.3	49.3	44.3	42.7	41.4	48.0	36.5	38.2	15.3	51.3	50.8	54.8	38.9	49.7
A12	F82	34.1	33.6	36.2	29.3	29.6	28.7	47.6	60.2	36.6	34.3	64.5	18.2	66.1	94.0	43.2	51.3
A13	F67	32.0	30.6	34.1	25.0	24.6	25.3	24.8	39.0	40.7	39.7	33.2	49.5	22.1	15.0	38.1	35.8
A14	F45	52.8	53.3	56.9	46.7	50.4	48.1	71.2	55.6	80.9	80.8	39.9	41.8	32.1	0.8	45.6	55.0
A15	F78	56.2	59.7	57.7	55.7	52.2	53.2	57.8	49.0	43.2	44.7	41.2	51.8	62.7	58.5	20.7	45.3
A16	F66	87.2	87.3	84.3	78.3	68.5	74.3	86.3	81.8	31.3	30.5	31.9	29.7	40.0	44.8	31.3	23.8