NECTEC-ACE 2026 ANNUAL CONFERENCE & EXHIBITION

The Rising Threat of Voice Spoofing

(เมื่อ AI ปลอมเสียงได้: ภัยคุกคามที่กำลังเพิ่มขึ้น)

งานประชุมวิชาการและนิทรรศการเนคเทค ประจำปี ๒๕๖๙

๓ กันยายน ๒๕๖๙

เจษฎา กาญจนะ

ศูนย์เทคโนโลยีอิเล็กทรอนิกส์และคอมพิวเตอร์แห่งชาติ

FAKE or REAL?

- How confident are you?

2019: A voice was enough.

- Fake CEO Voice → €220,000 stolen

WSJ PRO

Fraudsters Used AI to Mimic CEO's Voice in Unusual Cybercrime Case

Scams using artificial intelligence are a new challenge for companies

By Catherine Stupp

WSJ PRO Updated Aug. 30, 2019 at 12:52 pm ET

Forbes

A Voice Deepfake Was Used To Scam A CEO Out Of $243,000

By Jesse Damiani, Former Contributor. ⓘ Jesse Damiani covers AI, ClimateTech, and emerging media.

Published Sep 03, 2019, 04:42pm EDT, Updated Sep 03, 2019, 04:43pm EDT

Source: https://www.wsj.com/articles/fraudsters-use-ai-to-mimic-ceos-voice-in-unusual-cybercrime-case-11567157402

2024: Seeing and hearing is no longer believing.

- Deepfake executives → $25.6 million stolen

South China Morning Post

Hong Kong police | Hong Kong / Law and Crime

‘Everyone looked real’: multinational firm’s Hong Kong office loses HK$200 million after scammers stage deepfake video meeting

Source: https://www.scmp.com/news/hong-kong/law-and-crime/article/3250851/everyone-looked-real-multinational-firms-hong-kong-office-loses-hk200-million-after-scammers-stage

2024: A fake voice tried to influence an election.

- This time, the target wasn't a bank account. It was an election.

FCC FINES MAN BEHIND ELECTION INTERFERENCE SCHEME $6 MILLION FOR SENDING ILLEGAL ROBOCALLS THAT USED DEEPFAKE GENERATIVE AI TECHNOLOGY

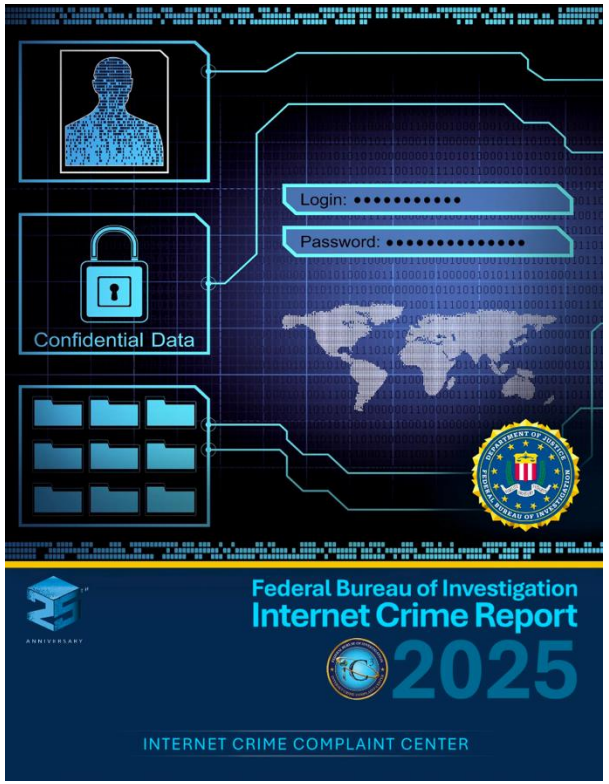
Steve Kramer Instigated Illegal Spoofed Robocall Campaign That Used a Deepfake of President Biden's Voice Telling Voters Not to Vote

WASHINGTON, September 26, 2024—The Federal Communications Commission today adopted a $6 million fine against political consultant Steve Kramer for illegal robocalls made using deepfake, AI-generated voice cloning technology and caller ID spoofing to spread election misinformation to potential New Hampshire voters prior to the state's January primary presidential election. Kramer will be required to pay the fine within 30 days or the matter will be promptly referred to the U.S. Department of Justice for collection.

Source: https://www.fcc.gov/document/fcc-issues-6m-fine-election-interference-deepfake-robocalls

2026: AI Voice Cloning Fuels Scams.

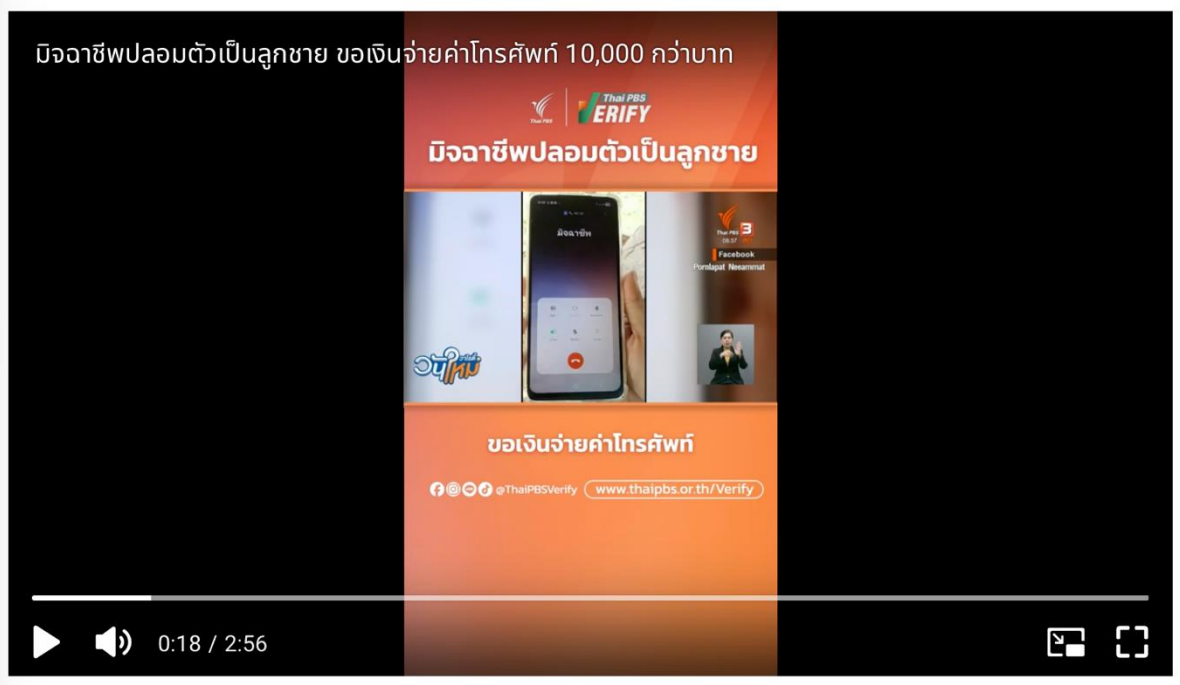
- U.S. Senator Maggie Hassan urged AI voice-cloning companies to strengthen safeguards against scams and fraud.
- This month (April), the FBI reported that scams involving AI, in general, accounted for more than $893 million in reported losses in 2025, and experts predict that generative AI tools could enable up to $40 billion in annual fraud losses in the United States by 2027.



Source: https://www.jec.senate.gov/public/index.cfm/democrats/2026/4/senator-hassan-presses-leading-ai-voice-cloning-companies-to-prevent-exploitation-by-scammers

2026: Voice Cloning Scams Reach Thailand

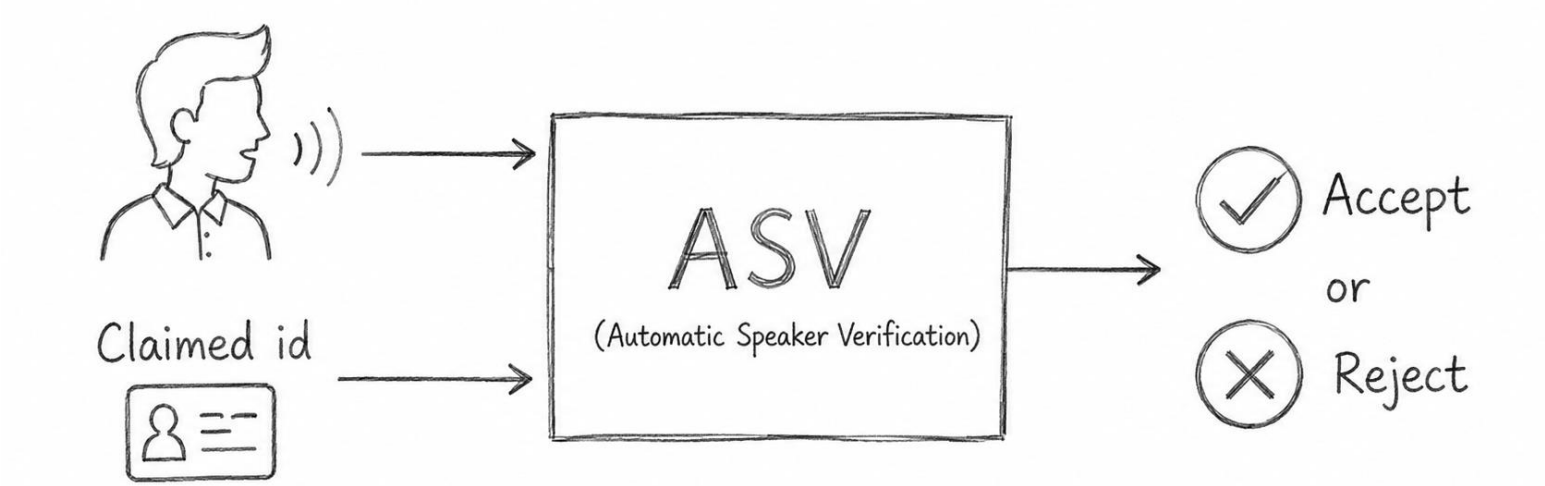
- "AI Clone & Deepfake" – Beware of 2026 New Scam Tactics as Voices and Faces of Your Family May Not Be Real (Thai PBS Verify)
- AI-based voice impersonation was already being used by scammers in Thailand in 2025.



Source: https://www.thaipbs.or.th/verify/en/article/content/8455

Voice is a biometric modality for identity verification.

Automatic Speaker Verification (ASV)



- ASV Applications
  - Access control
  - Forensic analysis
  - Voice-based authentication

HSBC Voice ID

- Voice biometrics for telephone banking authentication
- 2.8M+ active users
- voiceprint: My voice is my password.

Menu HSBC

4 May 2021

HSBC UK's Voice ID prevents £249 million of attempted fraud

HSBC UK has reported that telephone banking fraud is down 50% year-on-year.

HSBC UK's innovative voice biometrics system, Voice ID, has prevented almost £249 million of customers' money from falling into the hands of telephone fraudsters in the last year, with the rate of attempted fraud down 50% year-on-year, the bank announced today.

Source: https://www.about.hsbc.co.uk/news-and-media/hsbc-uks-voice-id-prevents-gbp249-million-of-attempted-fraud

NatWest

- Uses voice biometrics for customer authentication.

× **Can someone just use a recording of my voice?**

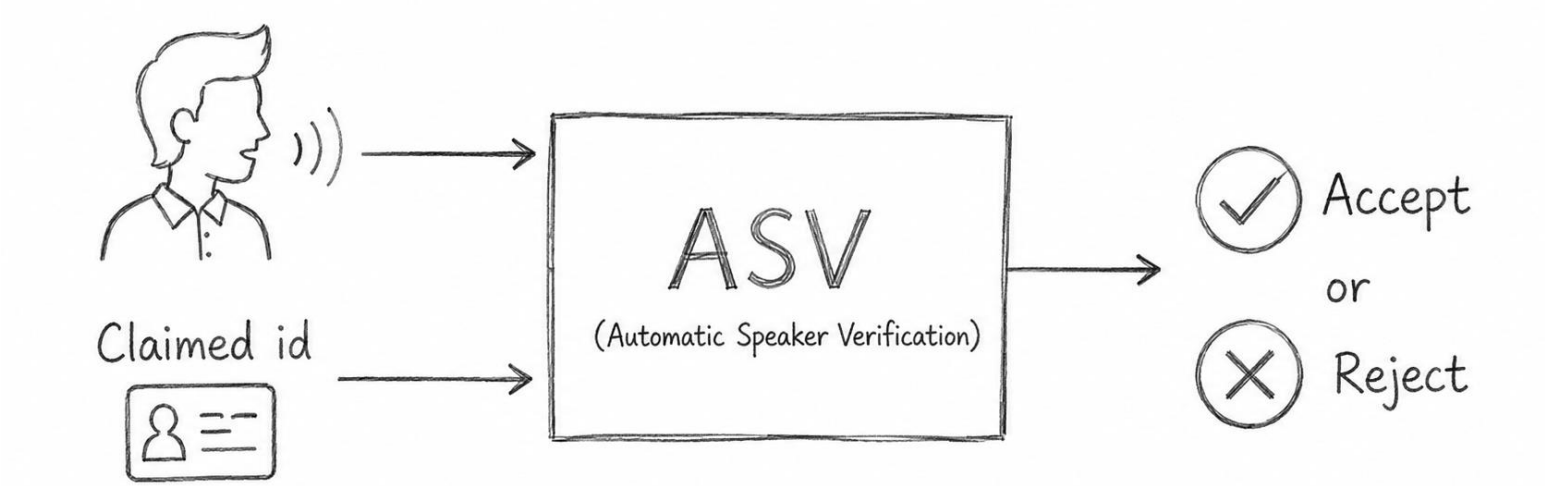
No. The biometric process is sophisticated, and considers many more things than just the sound of someone's voice. It will be able to tell if someone is playing back a recording, or even using artificial intelligence to mimic someone.

Source: https://www.natwest.com/support-centre/general-banking-information/general/what-is-biometric.html

Voice is a biometric modality for identity verification.

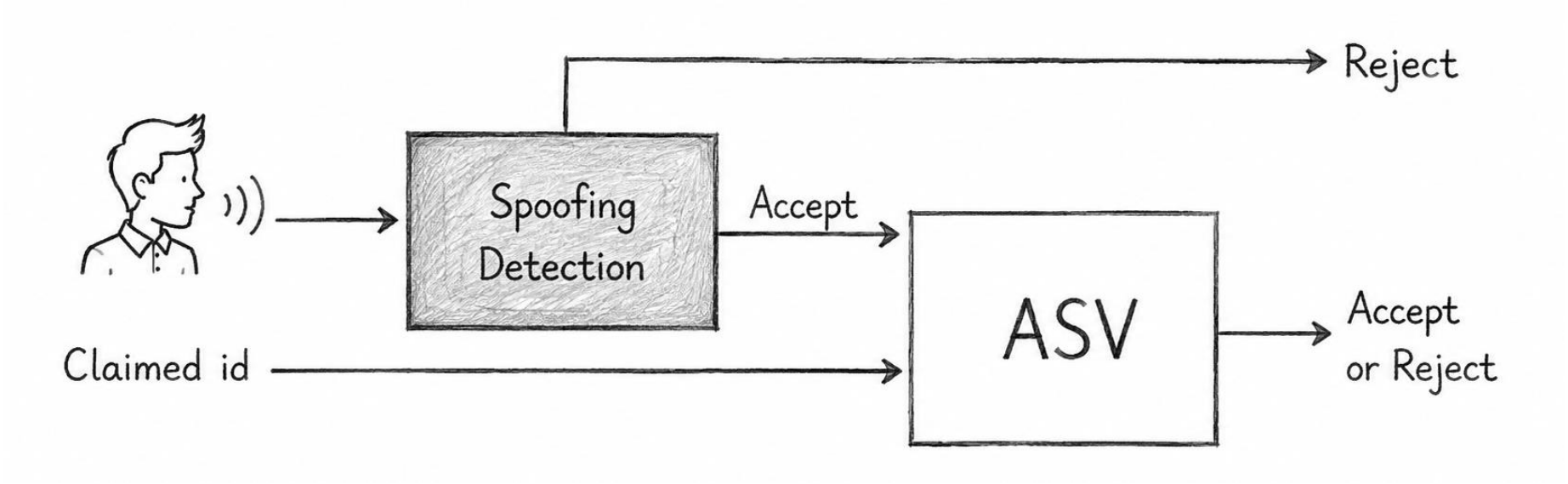
- What if a machine trusts your voice too?

Spoofing Attacks

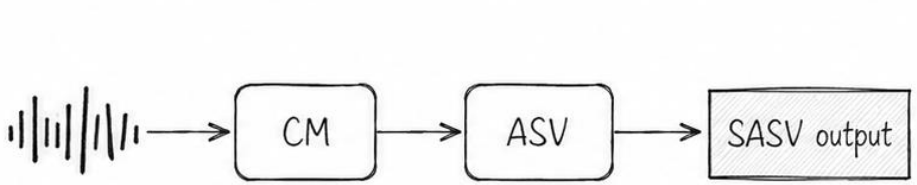


- ASV is vulnerable to spoof attacks.
  - Impersonation
  - Replay
  - Speech synthesis
  - Voice conversion

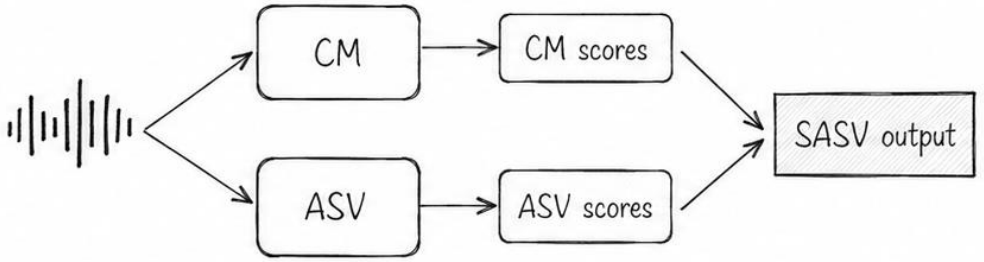
Countermeasure: Spoofing-aware ASV (SASV)



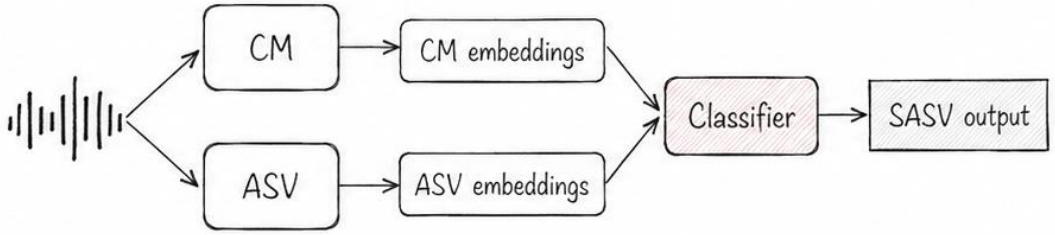
Different Structures of SASV



- Cascade system



- Score-level fusion



- Embedding-level fusion

- Integrated (E2E) system

Source: Menglu Li et al., Audio Anti-Spoofing Detection: A Survey, 2024; doi: https://doi.org/10.48550/arXiv.2404.13914

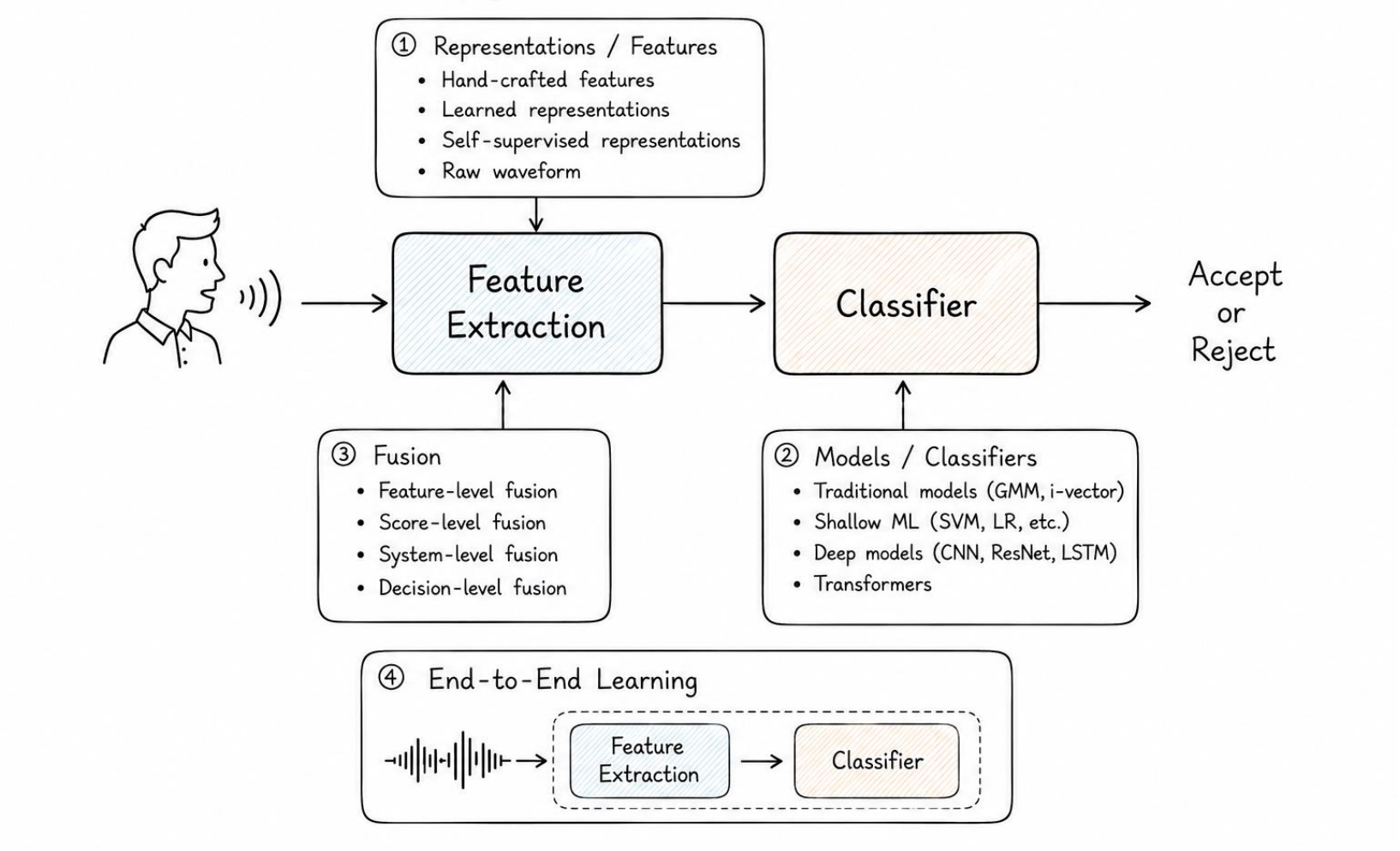
ASVspoof Challenge

- Launched in 2015 to advance research in automatic speaker verification spoofing and countermeasures
- Provides common datasets, protocols, and evaluation metrics
- Enables fair and reproducible comparison of different spoofing detection approaches
- Has become a major benchmark driving progress in speech anti-spoofing research

- Timeline: 2015 → 2017 → 2019 → 2021 → ASVspoof 5 (2024)

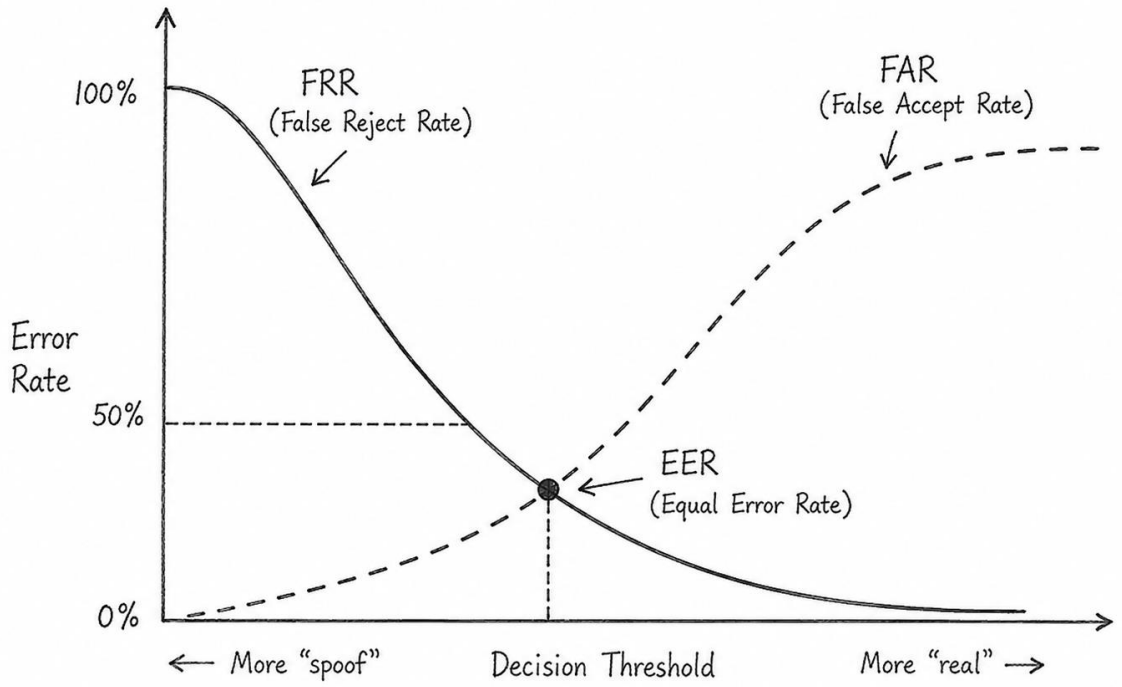
Source: https://www.asvspoof.org

Research Playground: Where Can We Improve?



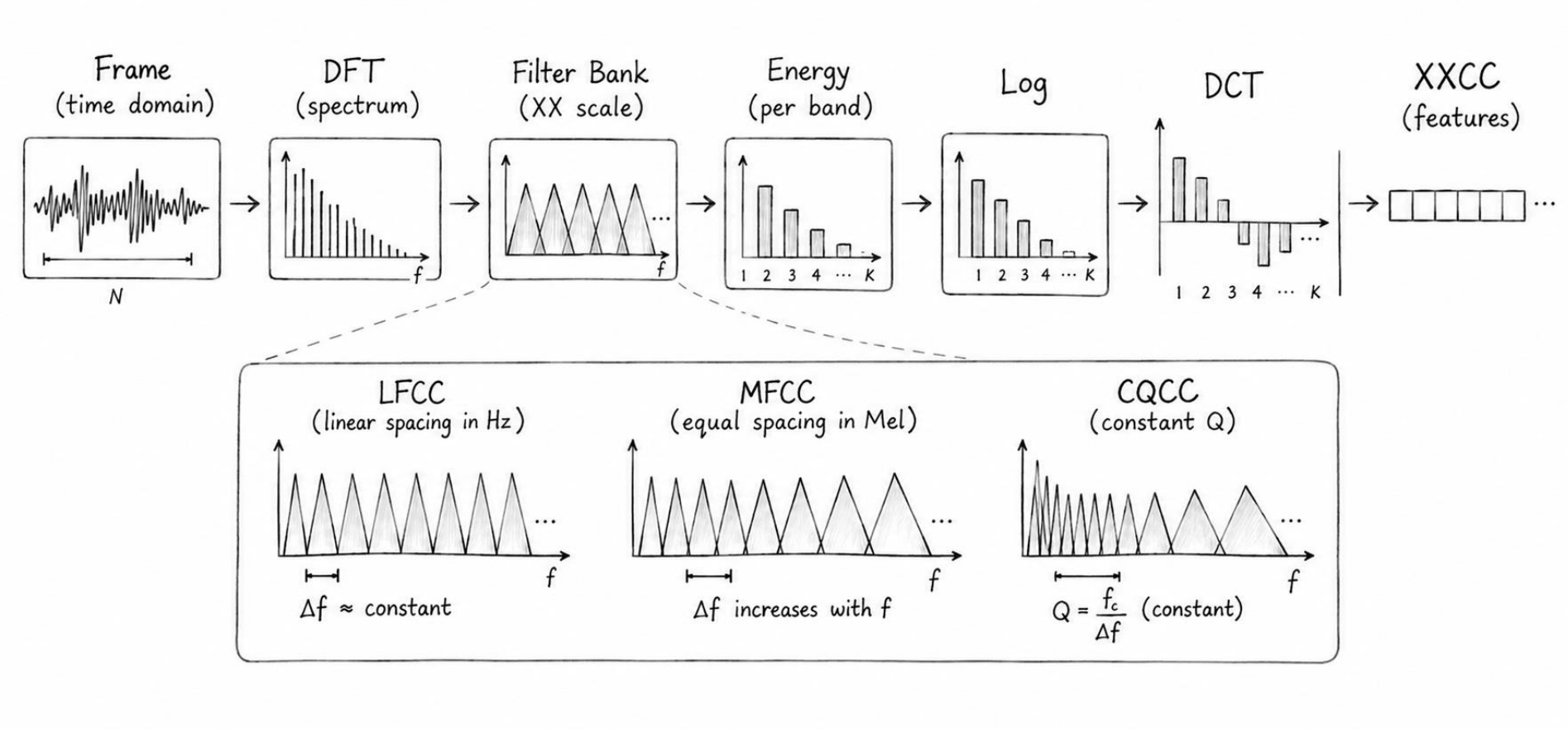
2016: CQCC

- Can we design features that make fake speech easier to distinguish from real speech?
- CQCC-GMM became a strong and widely used baseline for subsequent ASVspoof challenges.
- On **ASVspoof 2015**, CQCC-GMM achieved 0.46% EER.



Source: https://www.isca-archive.org/odyssey_2016/todisco16_odyssey.pdf

LFCC/MFCC/CQCC



2017: Deep Models & Fusion – Light CNN (LCNN)

- Can a deep model learn spoofing cues better than a conventional classifier?
- The top **ASVspoof 2017** system combined LCNN-based models with other classifiers through score fusion: 6.73% EER. (official baseline: 24.77% EER)

3.2. Convolutional neural network

We proposed the spoofing detection method based on CNN with Max-Feature-Map activation. MFM fuction is defined as

$$y_{ij}^{k} = max(x_{ij}^{k}, x_{ij}^{k+\frac{N}{2}}),$$
$$\forall i = \overline{1, H}, j = \overline{1, W}, k = \overline{1, N/2}$$

where x is the input tensor of size $H \times W \times N$, y is the output tensor of size $H \times W \times \frac{N}{2}$. Here i, j indicates the frequency and time domains and k is the channel index. Figure 3 illustrates MFM for a convolutional layer. MFM usage allowed us to reduce CNN architecture. That's why such CNN architecture is called Light CNN (LCNN) [16].

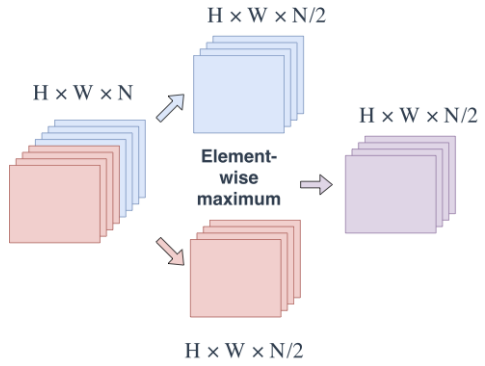
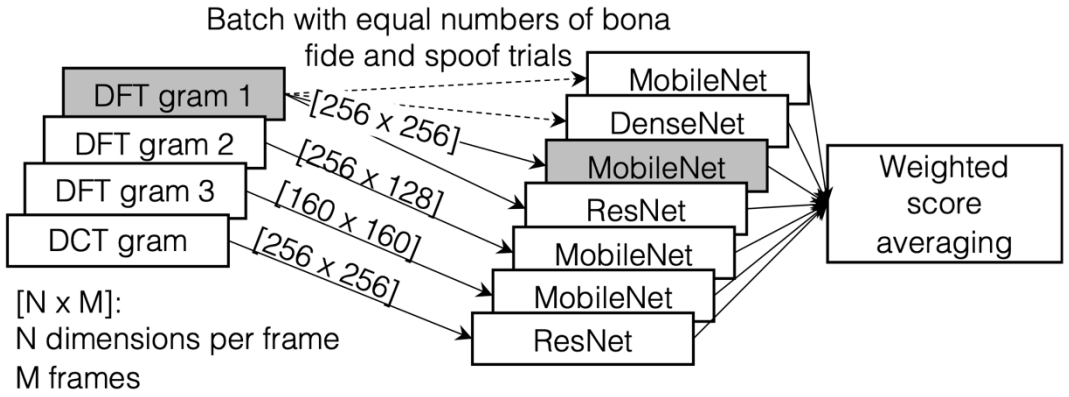


Figure 3: *MFM for convolutional layer*

Source: https://www.isca-archive.org/interspeech_2017/lavrentyeva17_interspeech.pdf

2019: Fusion

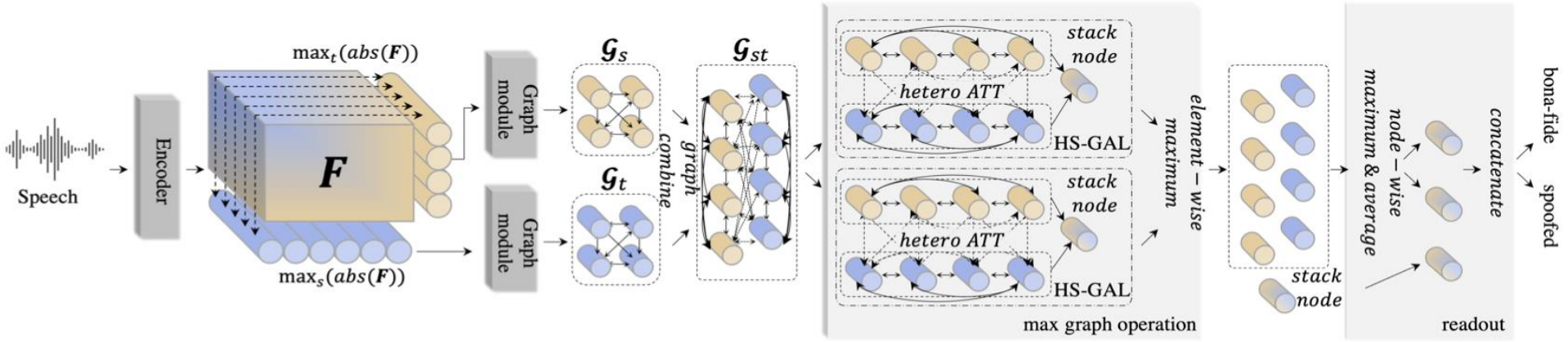
- Can complementary systems work better together?
- Different systems see different clues.
- **ASVspoof 2019** top system T05 fused seven sub-systems, using several spectral representations and classifiers.
- **ASVspoof 2019 LA**: 0.22% EER. (LFCC-GMM: 8.09%, CQCC-GMM: 9.57%)



Source:
https://www.researchgate.net/publication/349234464_ASVspoof_2019_spoofing_countermeasures_for_the_detection_of_synthesized_converted_and_replayed_speech

2020-2021: End-to-End

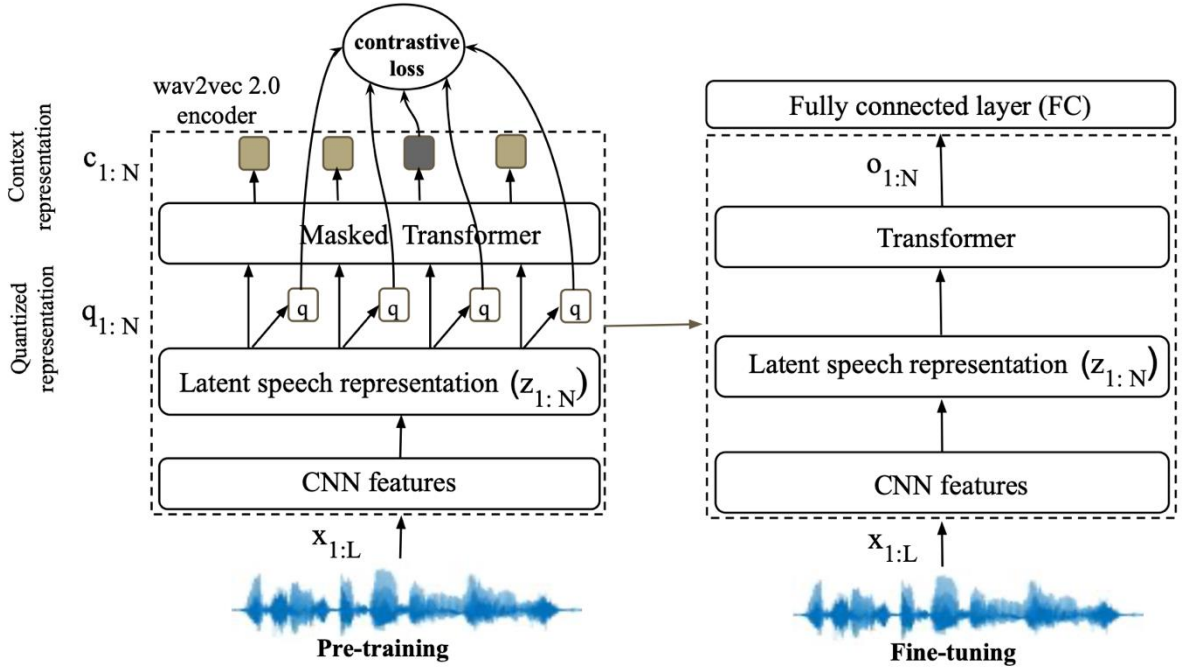
- Do we need to design acoustic features at all?
- RawNet2
- AASIST (Audio Anti-Spoofing using Integrated Spectro-Temporal Graph Attention Networks)
  - **ASVspoof 2019 LA**: 0.83% EER



Source: https://arxiv.org/pdf/2110.01200

2021-Present: Self-Supervised Learning

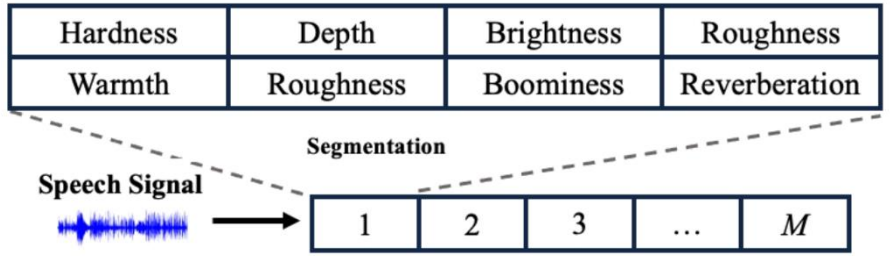
- What if the model learns speech representations before learning spoofing?
- Use pretrained speech models as front-ends (e.g., wav2vec 2.0, XLS-R, WavLM) and then fine-tune the pretrained model for spoofing detection
- wav2vec 2.0 + AASIST: 0.82% EER on **ASVspoof 2021 LA**

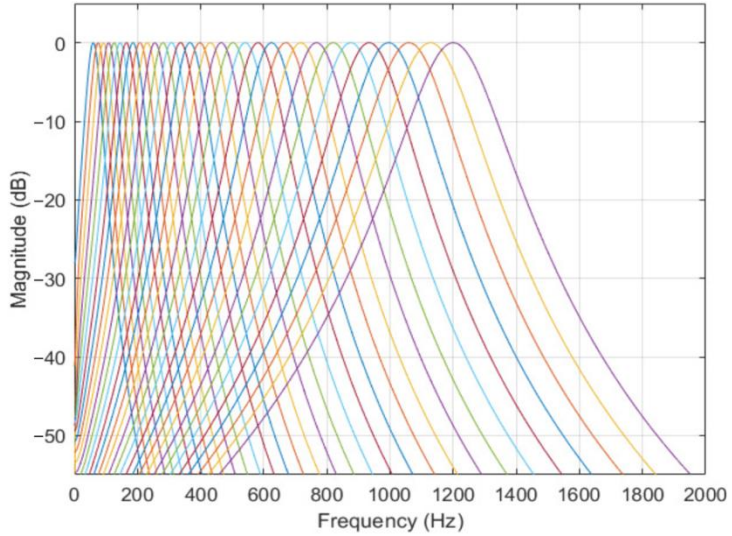


Source: https://arxiv.org/pdf/2202.12233

2026: Perceptual Pathological Features

- Can perceptual abnormalities in synthetic speech reveal deepfake?
- Explores speech-pathological / timbral attributes as deepfake cues
- Combines two complementary systems:
  - Pathological features → DNN
  - Gammatone filterbank → ResNet-18
- Evaluated on **ASVspoof 2019**: 5.93% EER

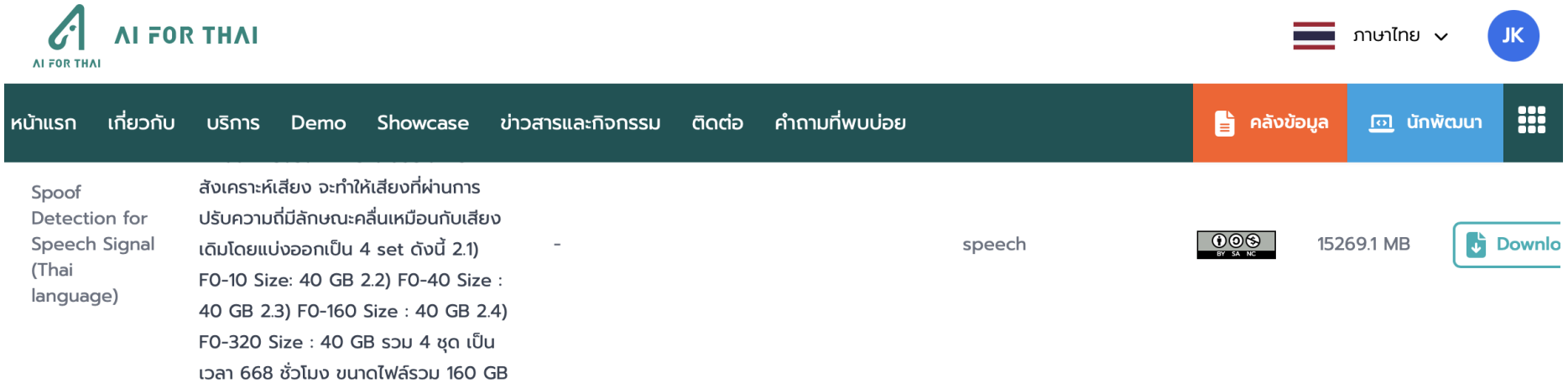




Source: https://www.mdpi.com/2076-3417/16/4/2077

ThaiSpoof (2023)

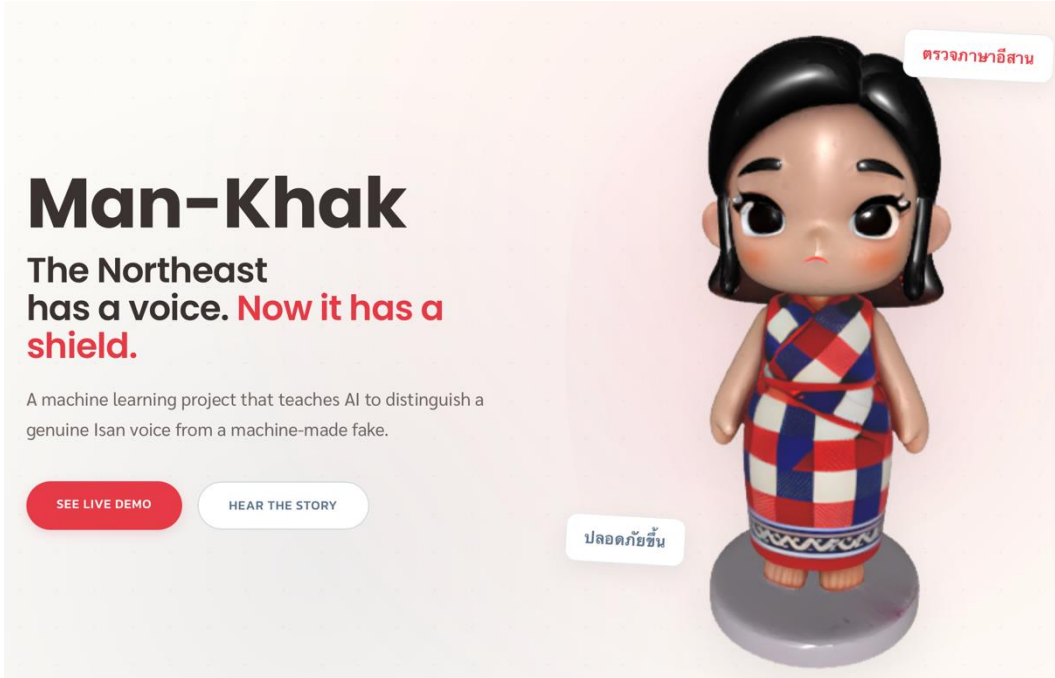
- Thai language dataset for spoofing detection



Source: AI for Thai

Isan Anti-Spoofing

- https://mankhak.vercel.app



Source: https://mankhak.vercel.app

Challenges

- Unknown generators
- Dataset dependency & cross-dataset generalization
- Transmission channels
- Noise & real acoustic environment
- Multilingual & cross-lingual speech
- Partial deepfakes
- Robustness to adversarial attacks

Takeaways

- Voice is a biometric, but it can be spoofed.
- Spoofing attacks are rapidly evolving.
- Countermeasures have evolved too.
- Excellent benchmark performance does not guarantee real-world robustness.
- Generalization remains a key challenge.

The goal is not just to detect the attacks we know, but to detect the attacks we have never seen.