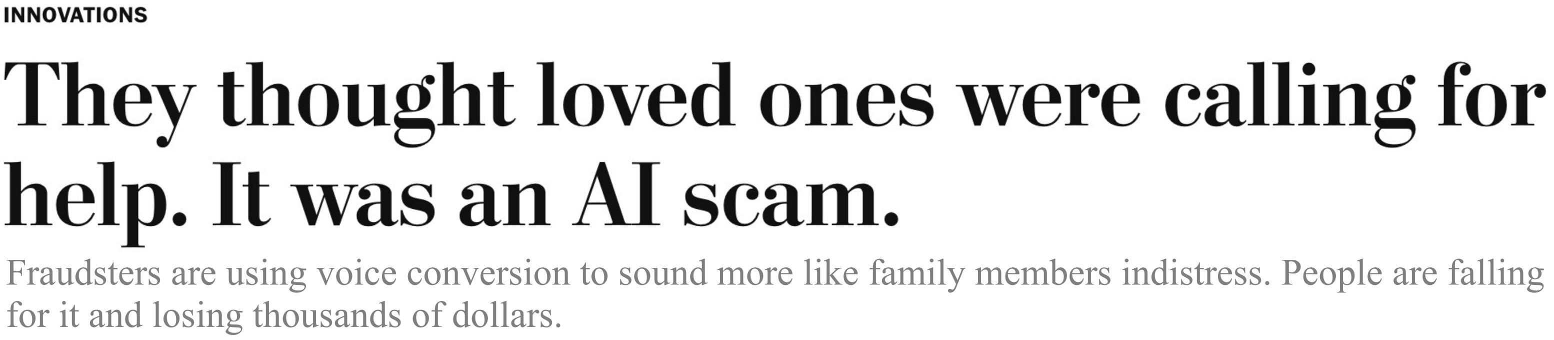
REVELIO: REvealing Source VoicEprint ConceaLed by VoIce COnversion

Jiangyi Deng[1], Yanjiao Chen*[1], Yinan Zhong[1], Qiaohao Miao[1], Xueluan Gong[2], Wenyuan Xu[1]

[1]Ubiquitous System Security Lab (USSLAB), Zhejiang University

[2]Wuhan University

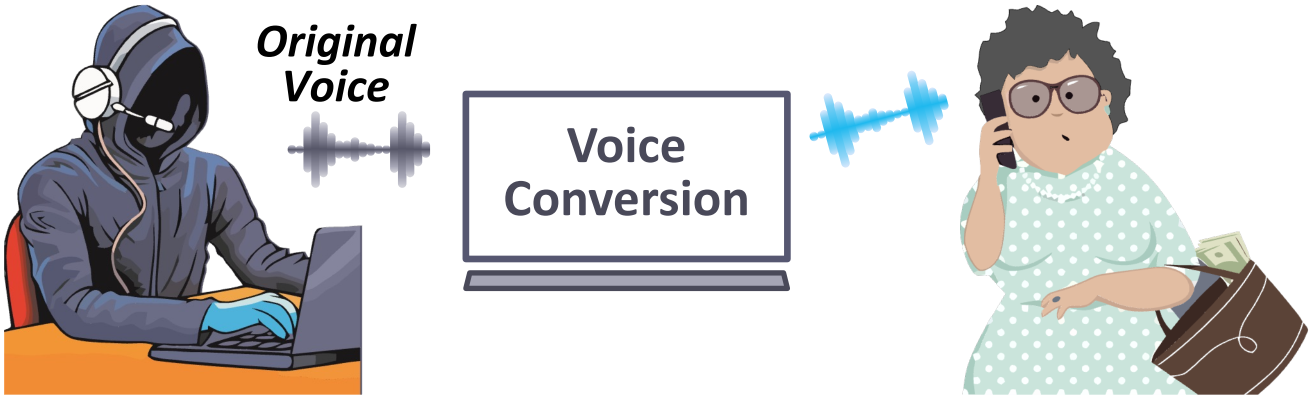
{jydeng, chenyanjiao, ynzhong, qhmiao, wyxu}@zju.edu.cn



INNOVATIONS

They thought loved ones were calling for help. It was an AI scam.

Fraudsters are using voice conversion to sound more like family members indistress. People are falling for it and losing thousands of dollars.

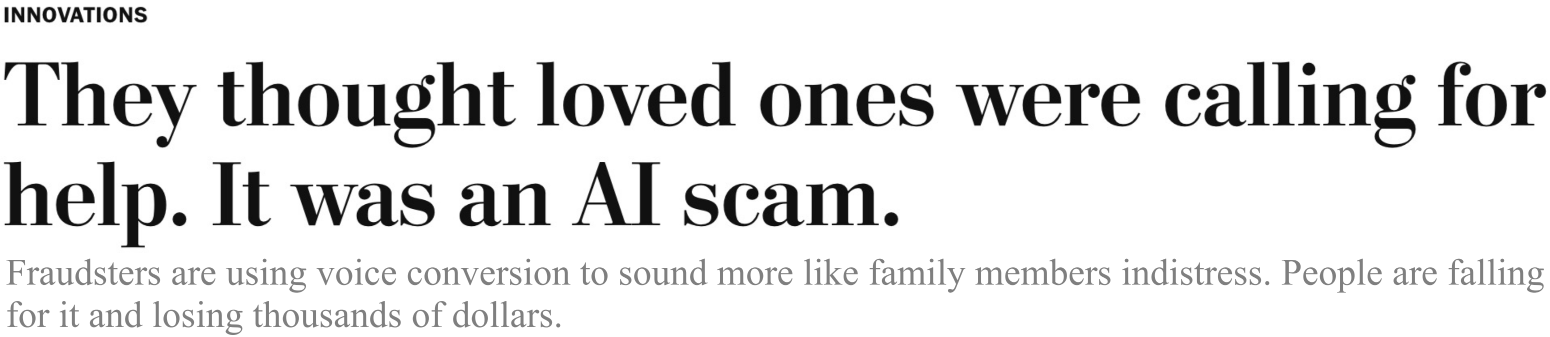


Original Voice

Voice Conversion

Fraudster

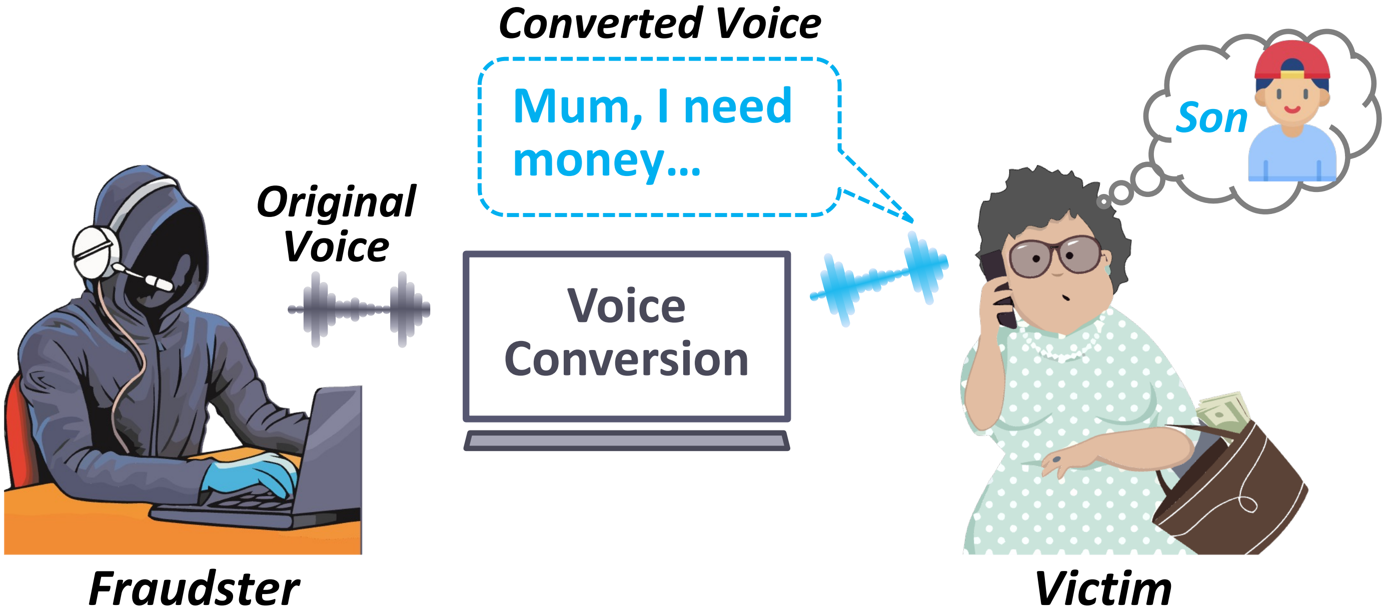
Victim



INNOVATIONS

They thought loved ones were calling for help. It was an AI scam.

Fraudsters are using voice conversion to sound more like family members indistress. People are falling for it and losing thousands of dollars.



Converted Voice

Mum, I need money…

Original Voice

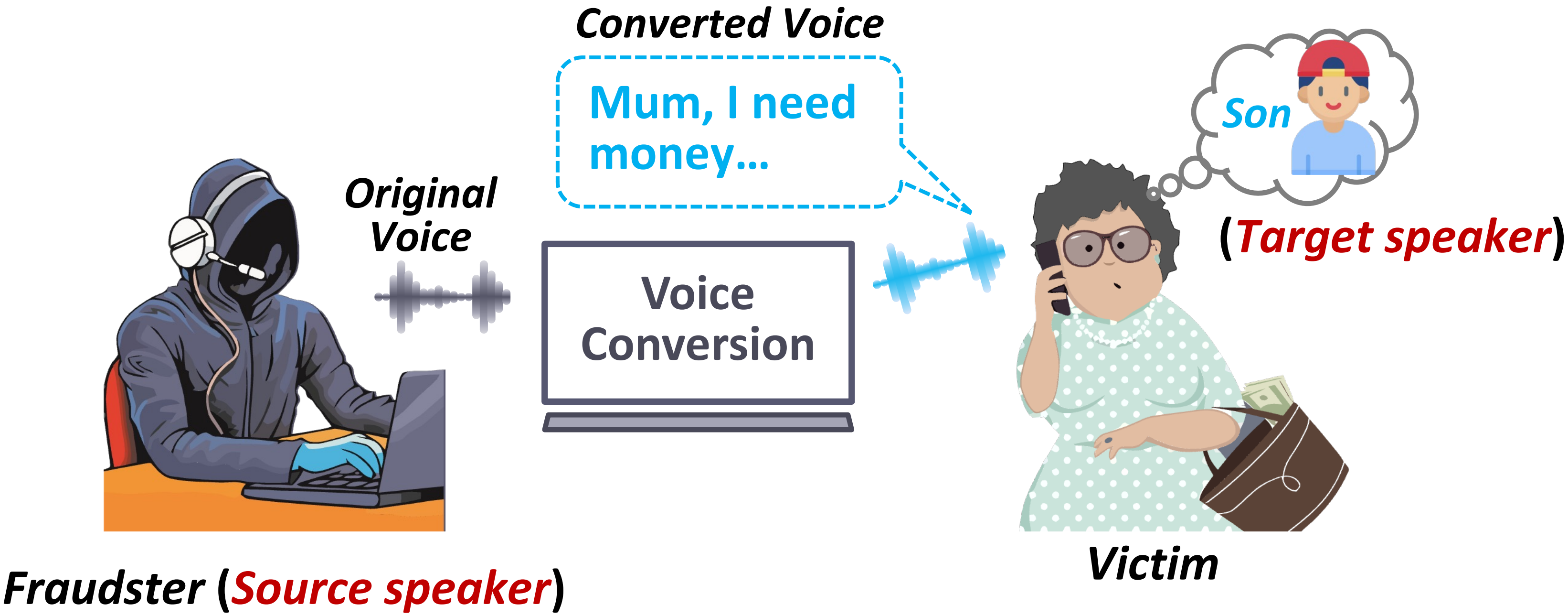
Voice Conversion

Son

Fraudster

Victim

Problem: New Phone Scams



Converted Voice

Mum, I need money…

Original Voice

Voice Conversion

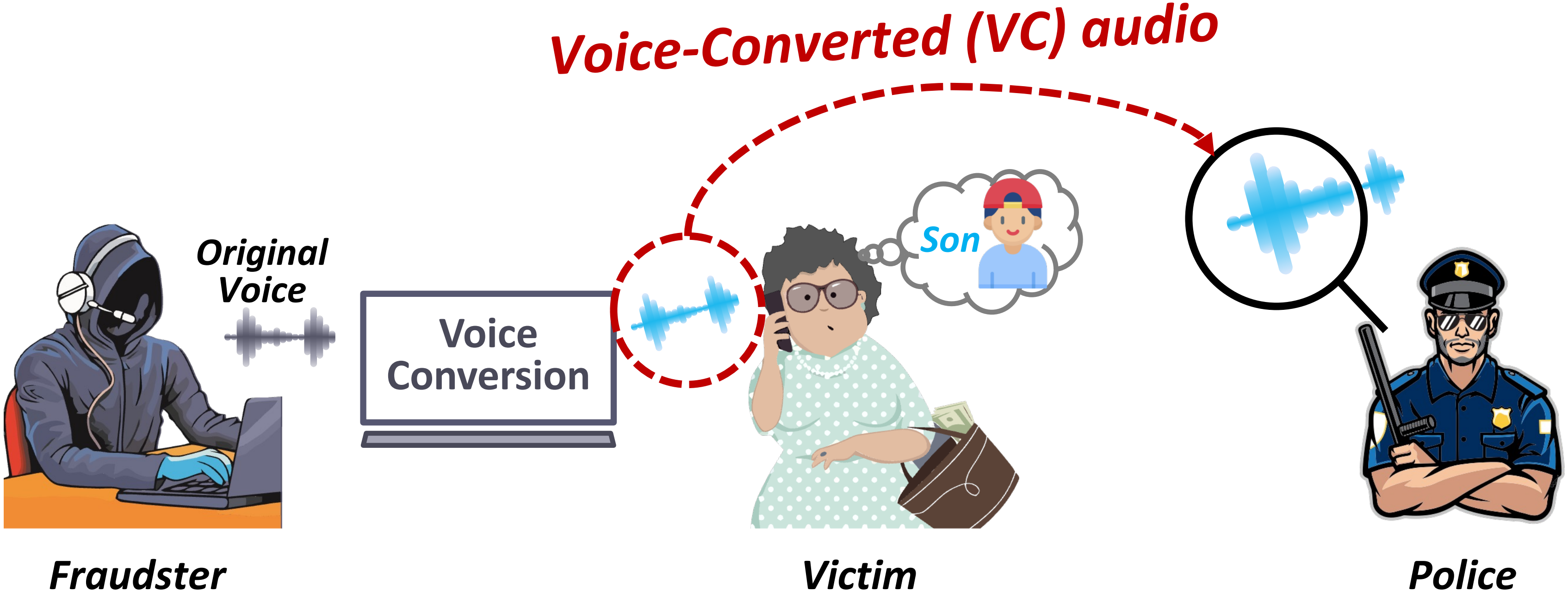
Son

(*Target speaker*)

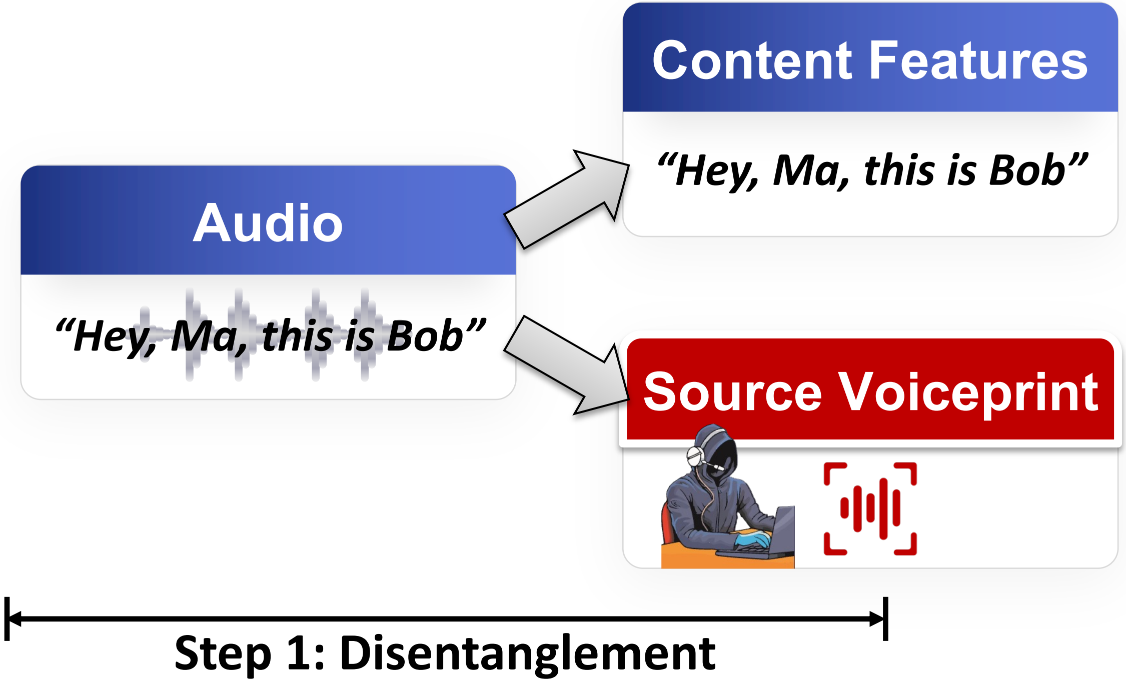
Fraudster (*Source speaker*)

Victim

Can We Identify the Fraudster from VC audio?



Is It Feasible to Identify the Source Speaker?



Audio

"Hey, Ma, this is Bob"

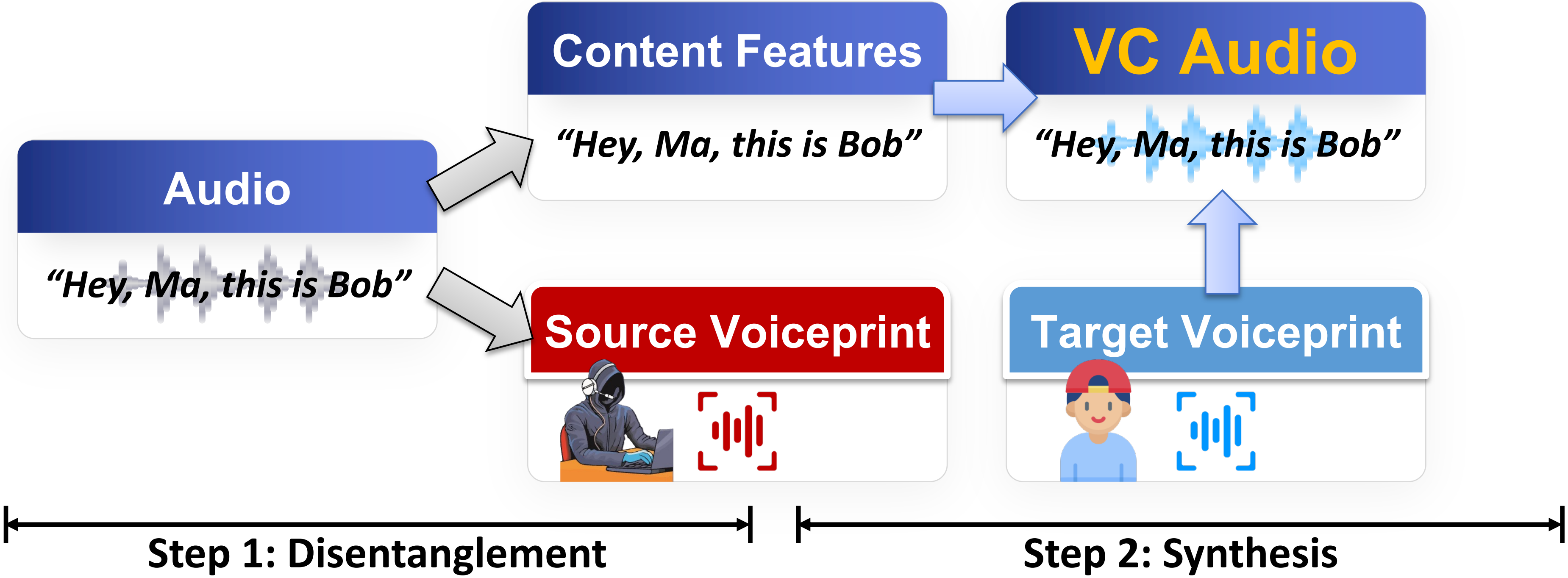
Content Features

"Hey, Ma, this is Bob"

Source Voiceprint

Step 1: Disentanglement

Is It Feasible to Identify the Source Speaker?



Audio

“Hey, Ma, this is Bob”

Content Features

“Hey, Ma, this is Bob”

VC Audio

“Hey, Ma, this is Bob”

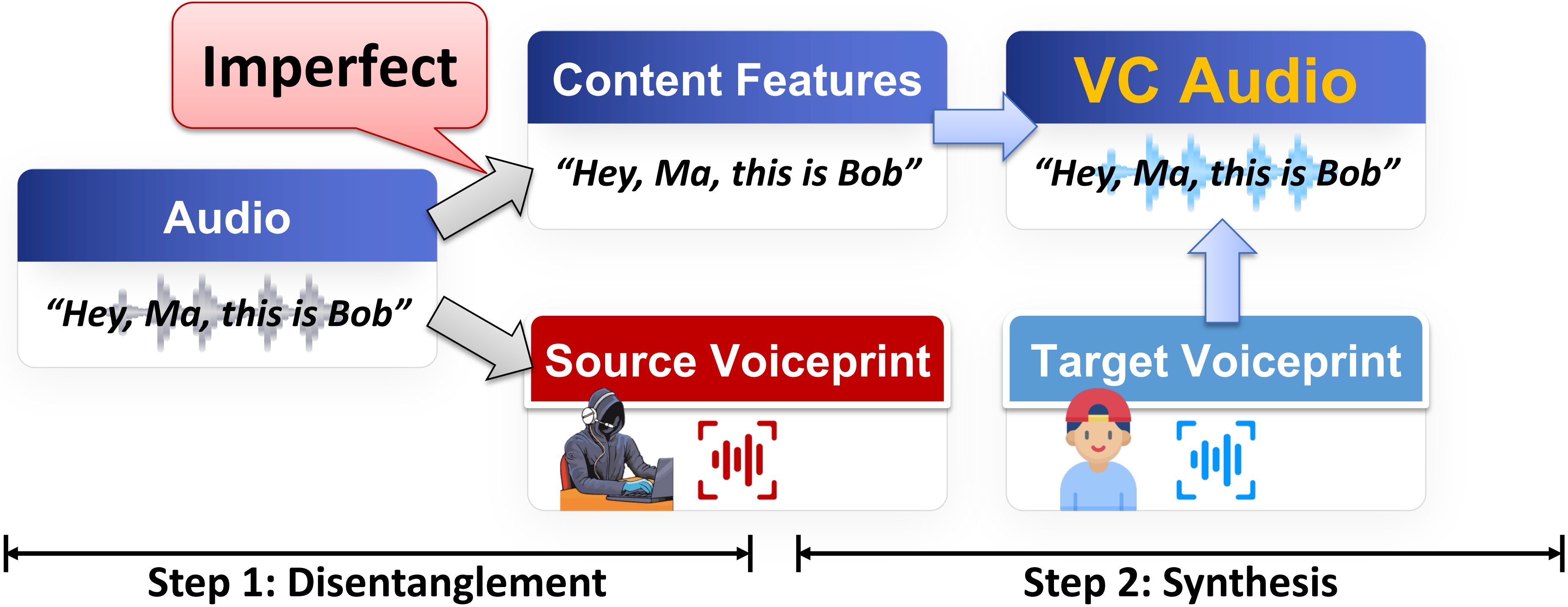
Source Voiceprint

Target Voiceprint

Step 1: Disentanglement

Step 2: Synthesis

Is It Feasible to Identify the Source Speaker?



Imperfect

Audio

"Hey, Ma, this is Bob"

Content Features

"Hey, Ma, this is Bob"

Source Voiceprint

VC Audio

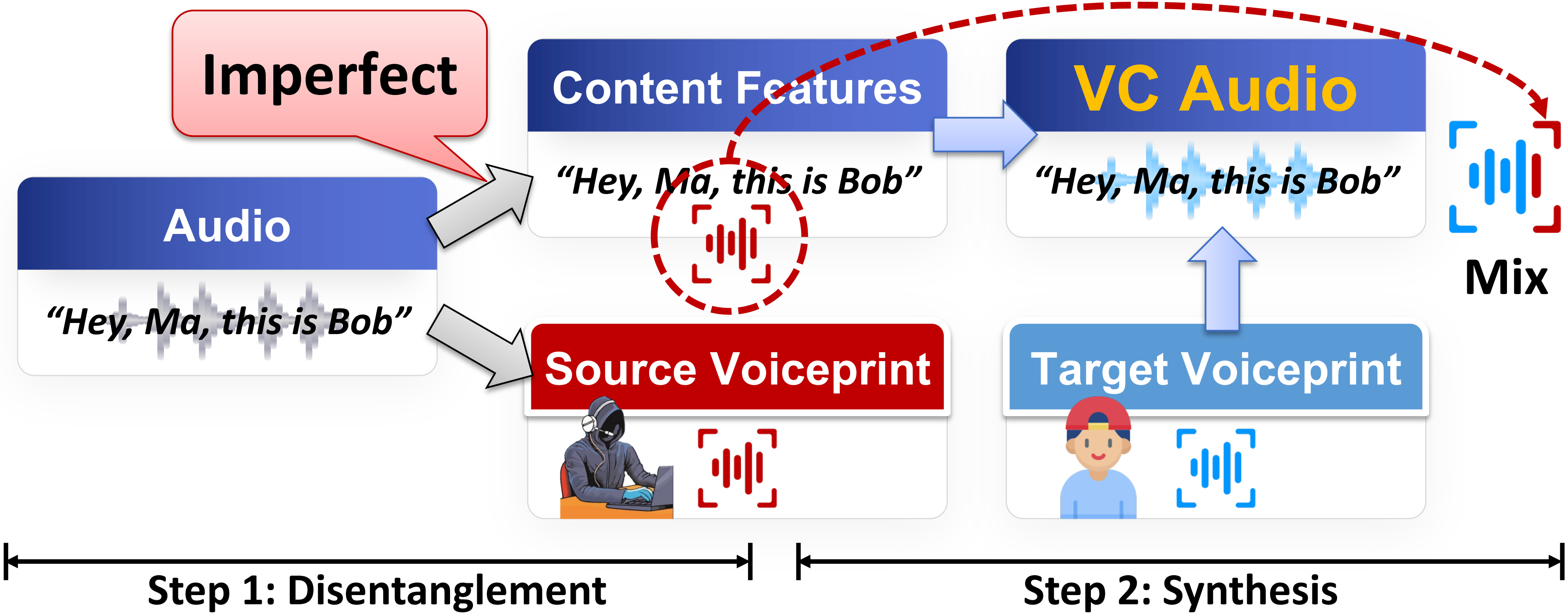
"Hey, Ma, this is Bob"

Target Voiceprint

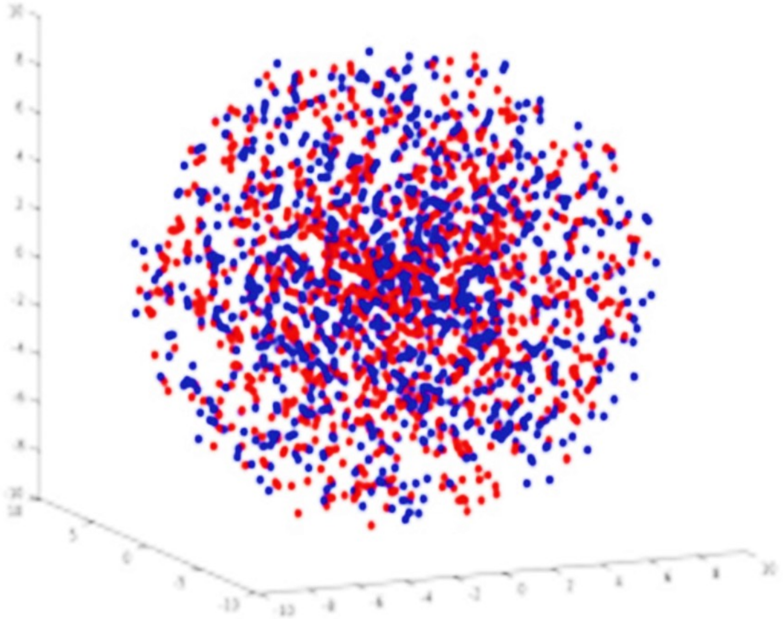
Step 1: Disentanglement

Step 2: Synthesis

Is It Feasible to Identify the Source Speaker?

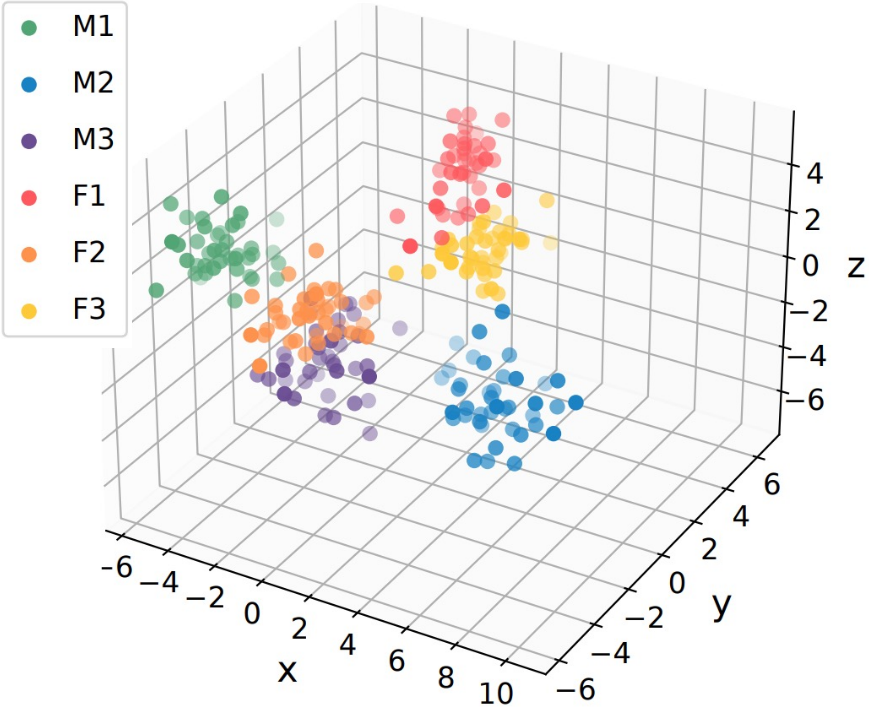


Ideal Disentanglement



- **Any fraudsters**
- **Content**: “Hello world”
- **Content features**: **perfectly overlapped**.

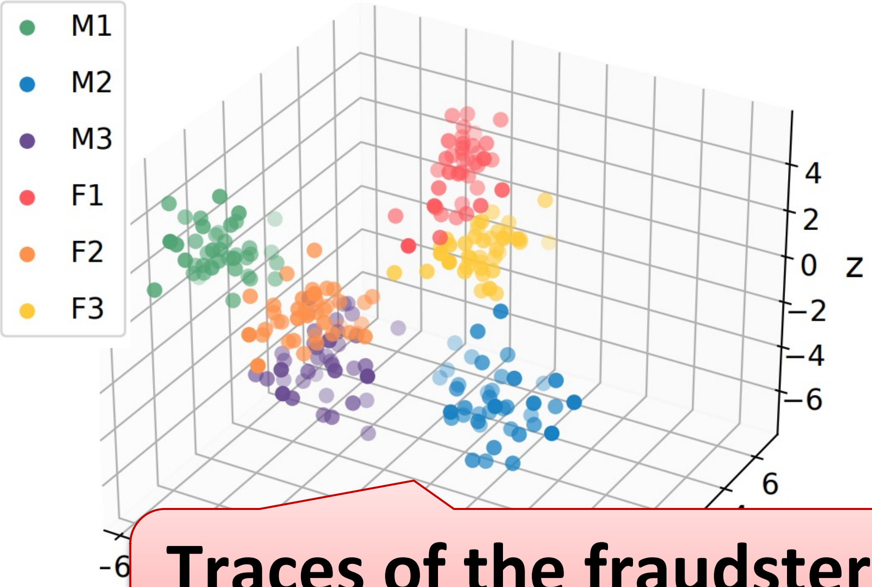
Imperfect Disentanglement Validation



Content Features of Six Fraudsters

- **Fraudster:** 3 males and 3 females
- **Content**: “Hello world”
- **Content features**: **not perfectly overlapped**.

Imperfect Disentanglement Validation

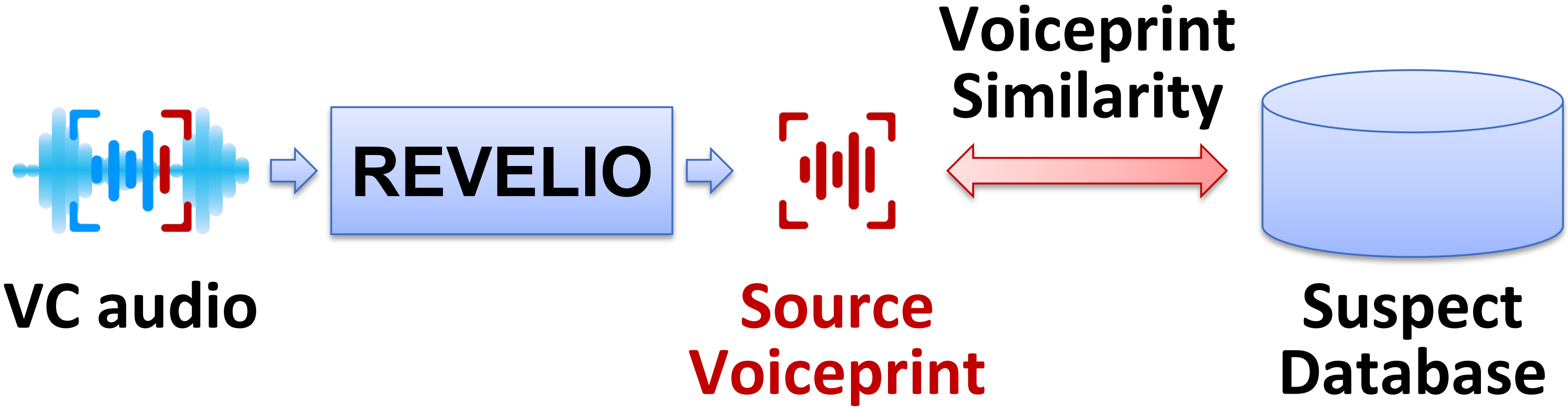


Traces of the fraudster voiceprint exist.

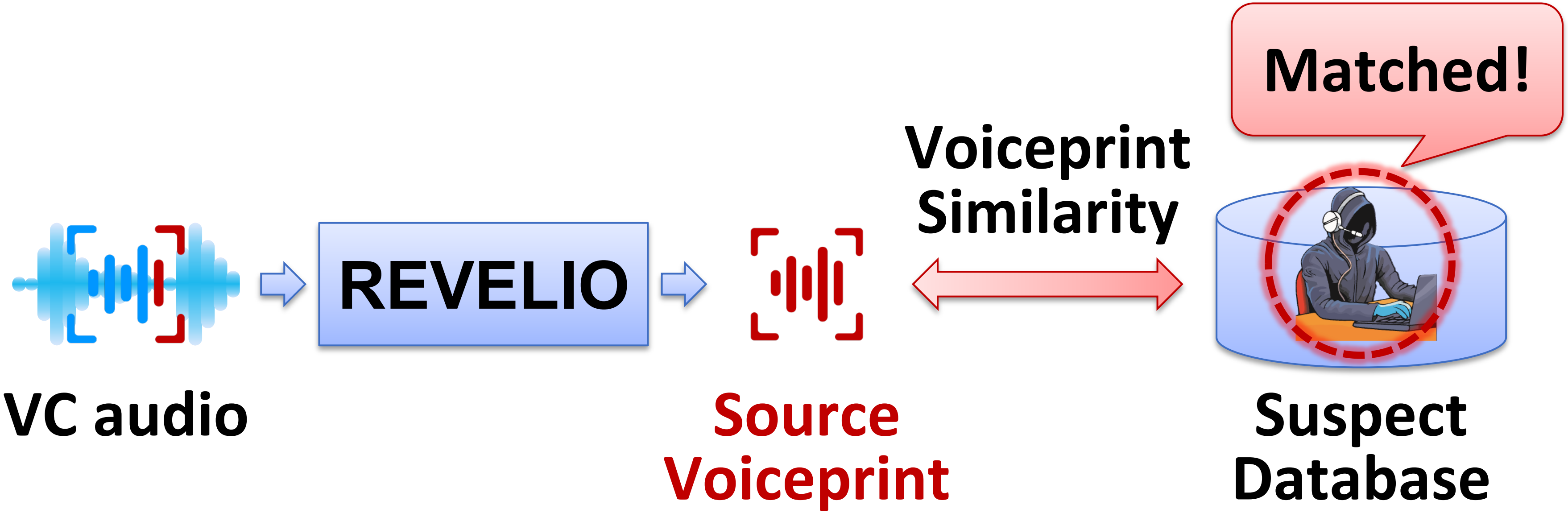
Content Features of Six Fraudsters

- **Fraudster:** 3 males and 3 females
- **Content**: “Hello world”
- **Content features**: **not perfectly overlapped**.

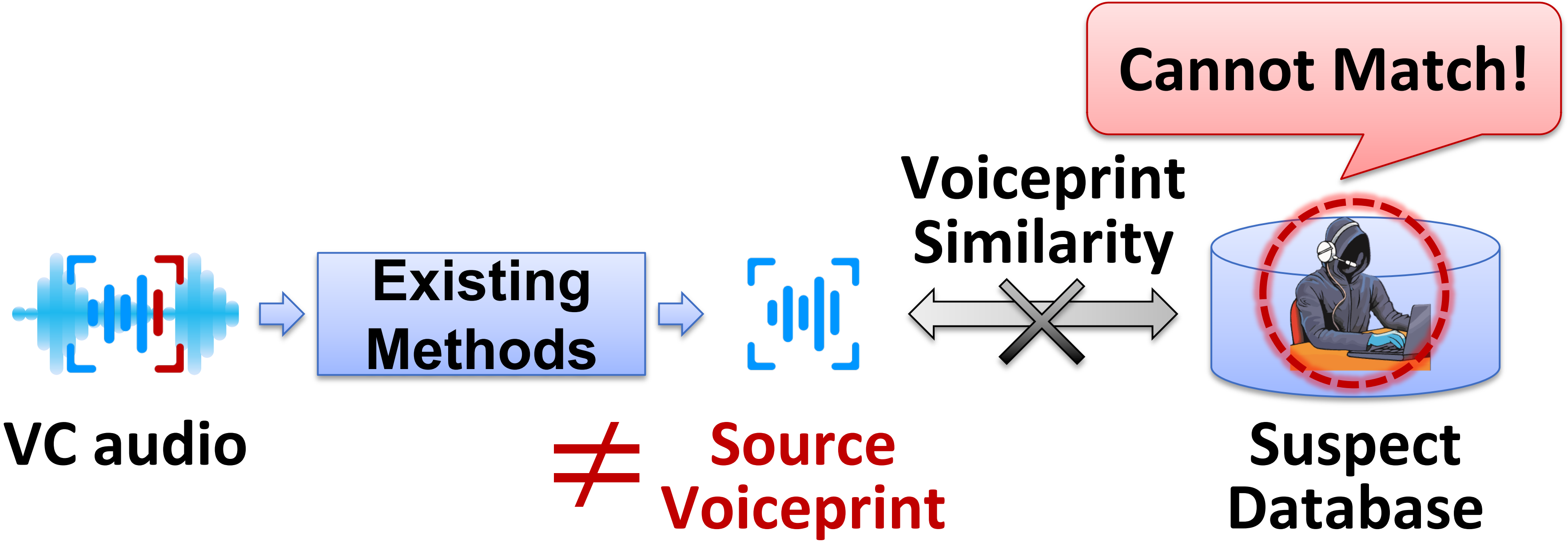
Threat Model



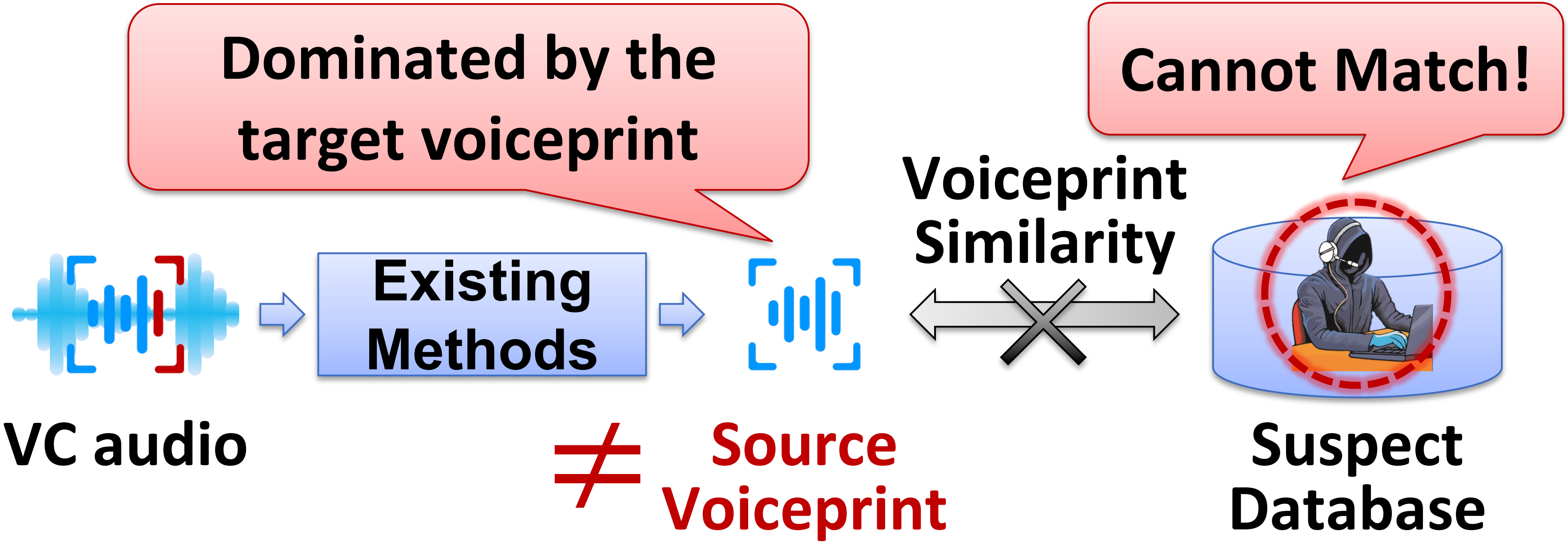
Threat Model



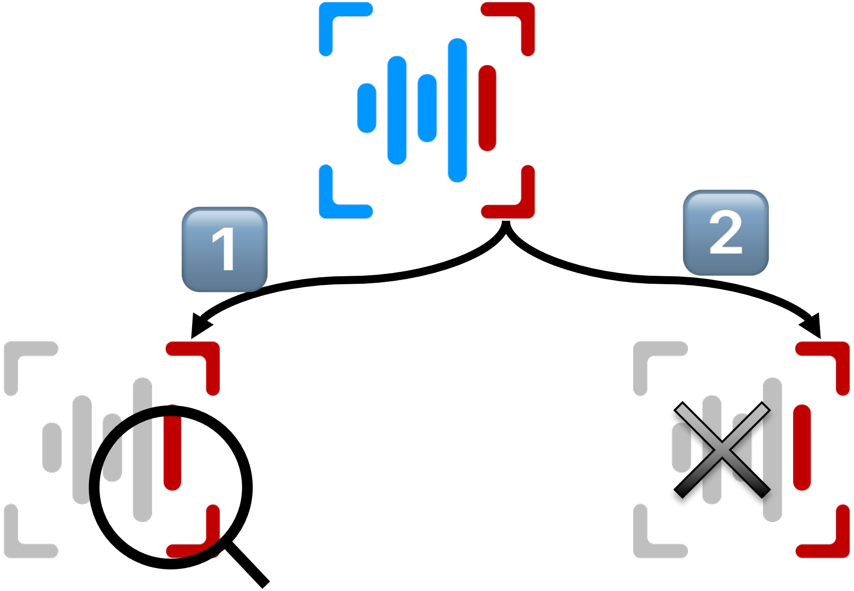
But in Reality …



But in Reality …



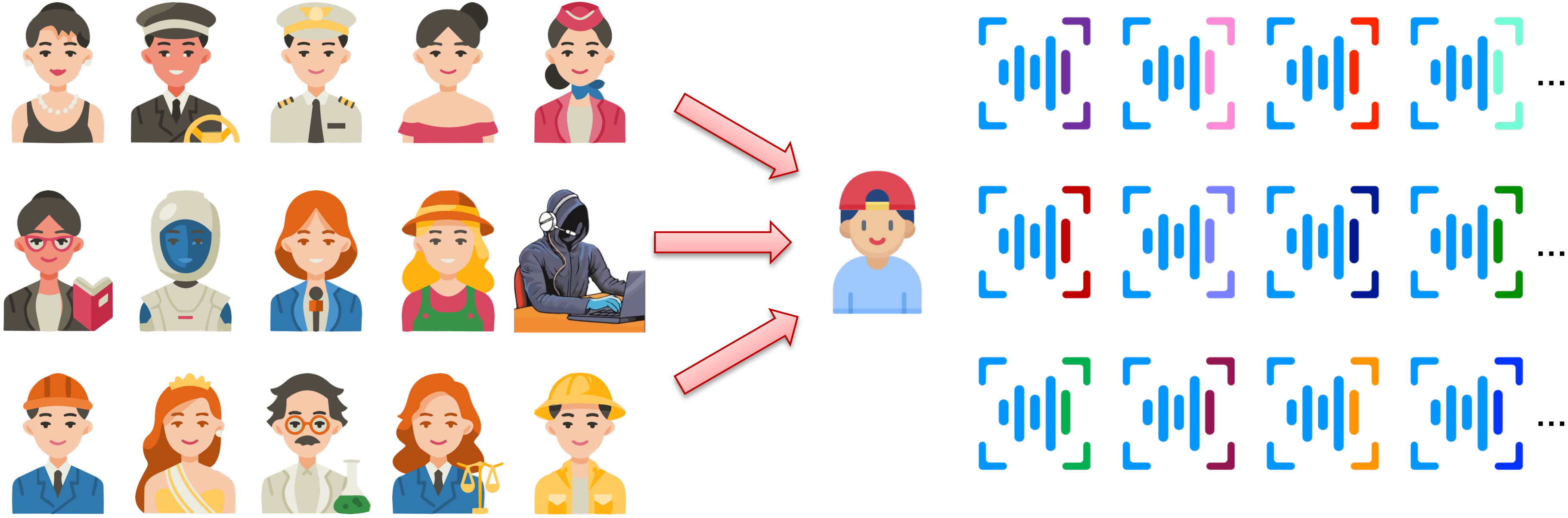
How to Extract the Source Voiceprint?



Extract the fraudster's features

Eliminate the target's features

To Better Extract the Fraudster’s Features



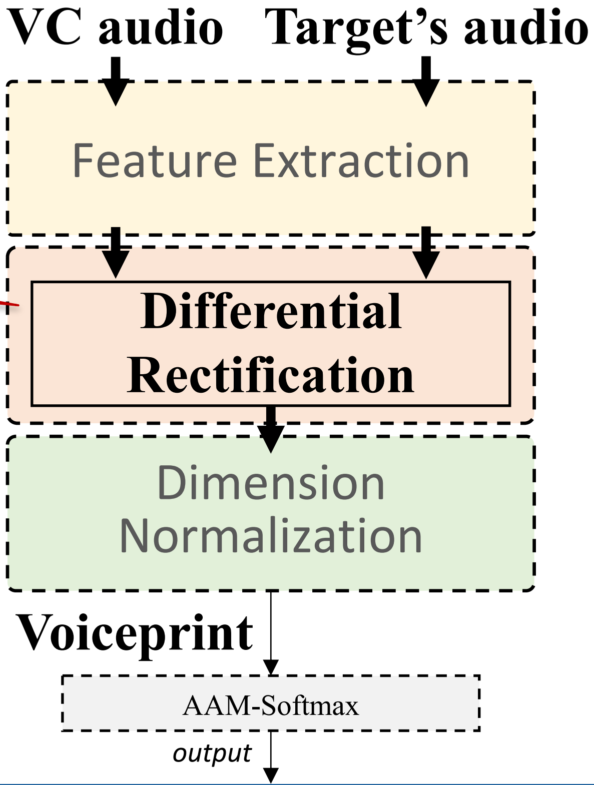
≈ 10, 000 source speakers

Differential Rectification

To eliminate the influence from target speaker

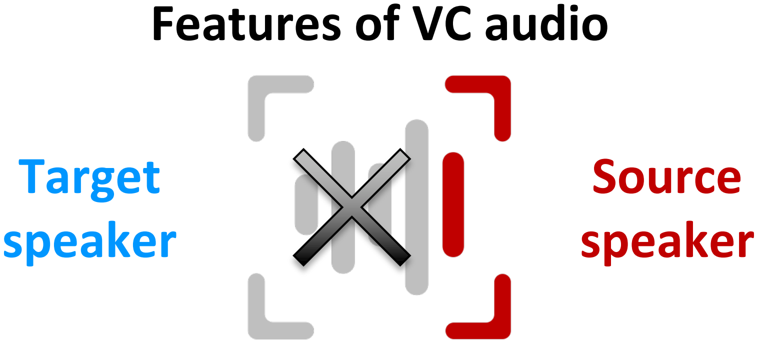
Features of VC audio

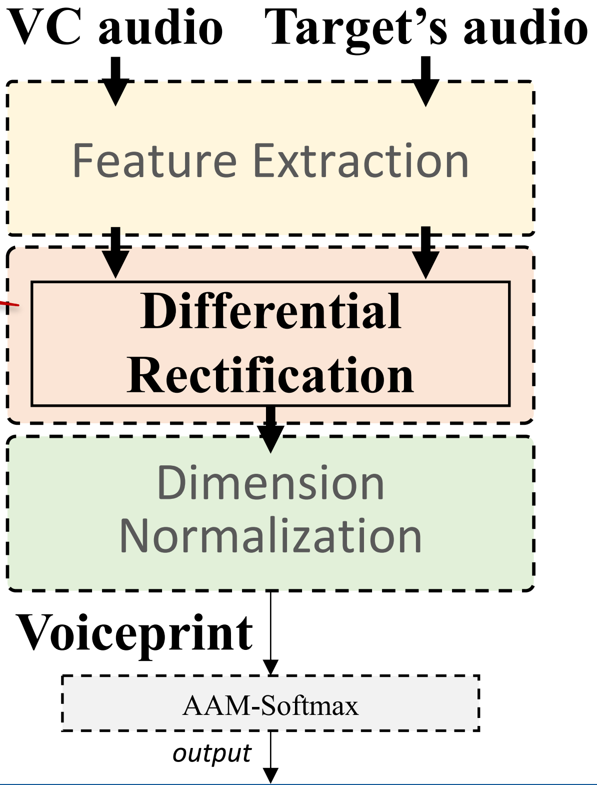
Target speaker
Source speaker



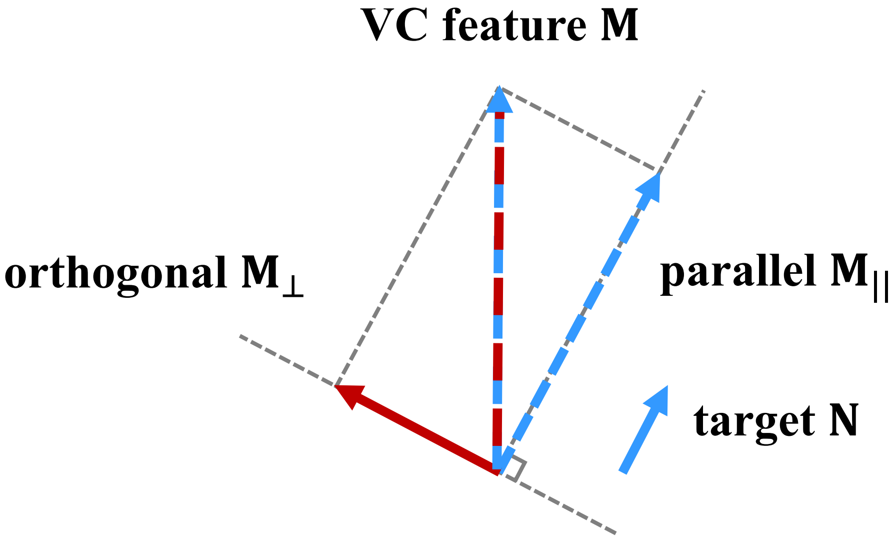
Differential Rectification

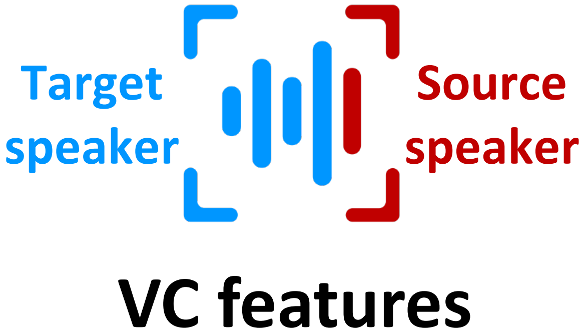
To eliminate the influence from target speaker



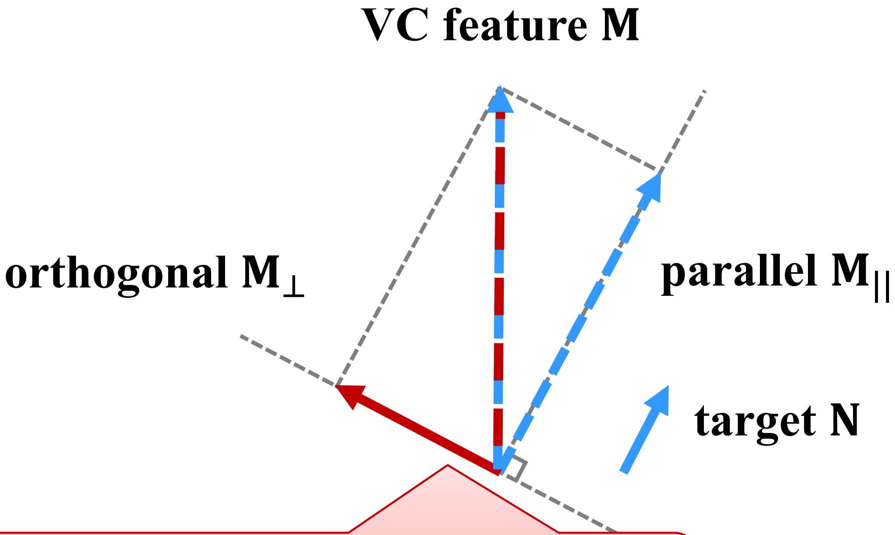


Differential Rectification





Differential Rectification



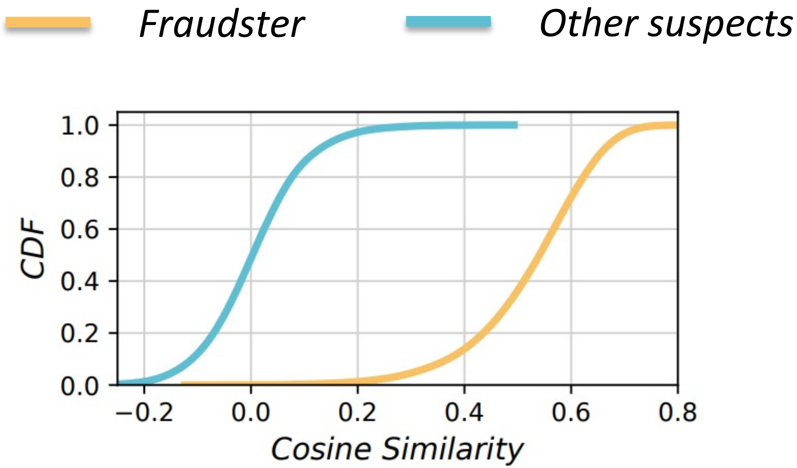
Represent the differences between the fraudster and the target

VC features

Experimental Setup

- **Voice conversion**
  - VQVC, VQVC+, AGAIN, BNE
- **Existing voiceprint extractors**
  - ECAPA-TDNN: SOTA speaker recognition model
  - Wang's: countermeasure for voice transformations
  - Zheng's: countermeasure for voice disguises

Effectiveness: Verify the Fraudster

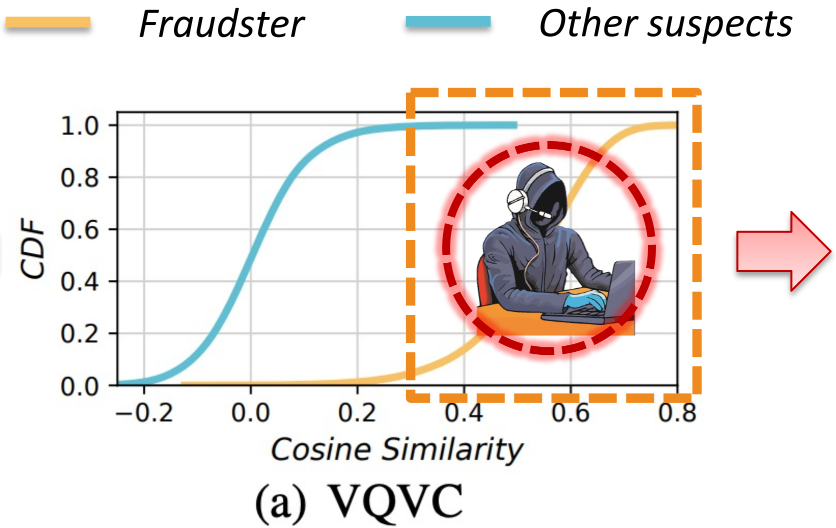


(a) VQVC

Voiceprint similarity between suspects and the VC audio.

Effectiveness: Verify the Fraudster

Low similarity
Innocent

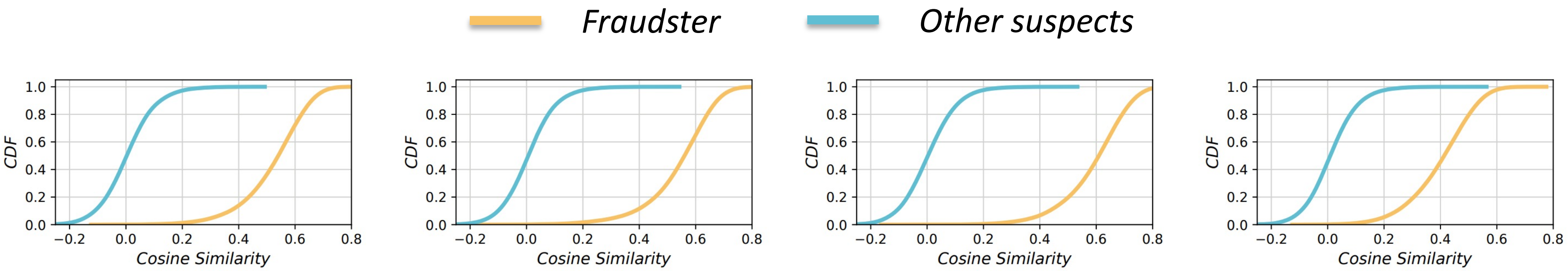


(a) VQVC

High similarity
Fraudster

The higher the similarity, the more likely it is to be the fraudster.

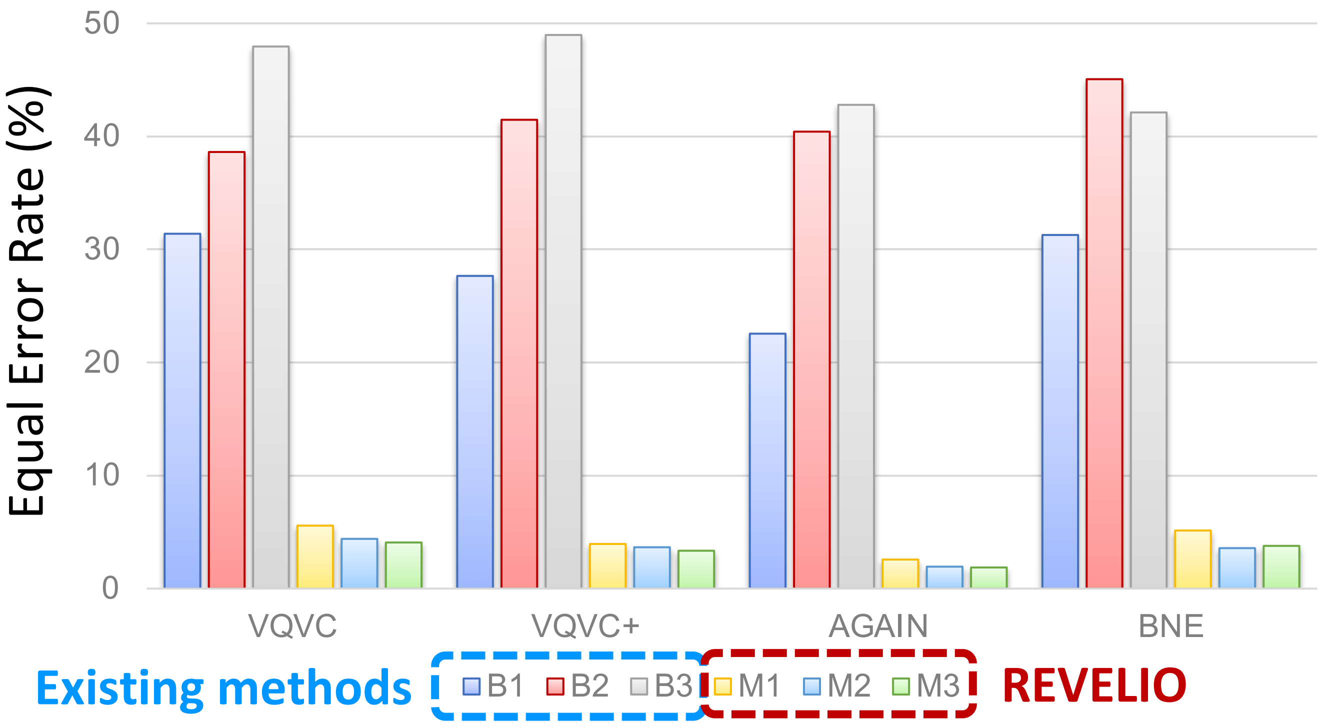
Effectiveness vs. VC Methods



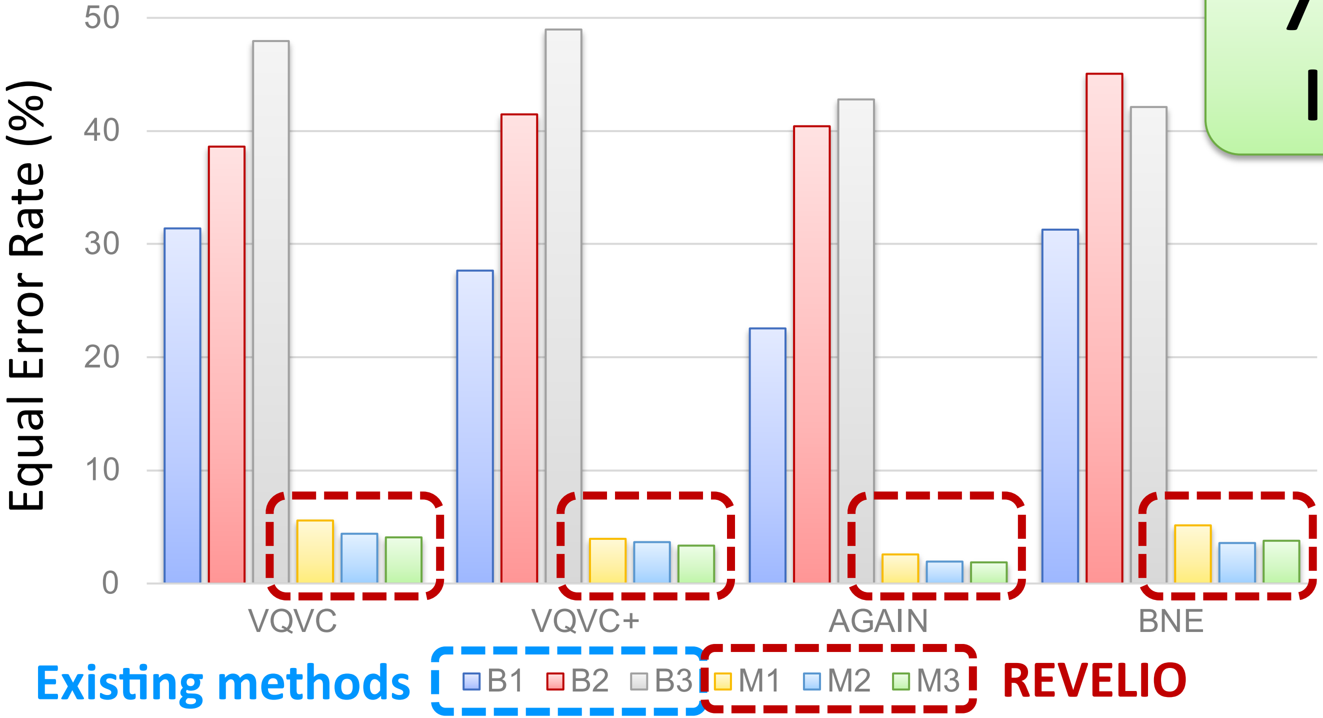
(a) VQVC (b) VQVC+ (c) AGAIN (d) BNE

REVELIO works for ***all 4 voice conversion*** *methods in our experiments.*

Compared with the Existing Methods

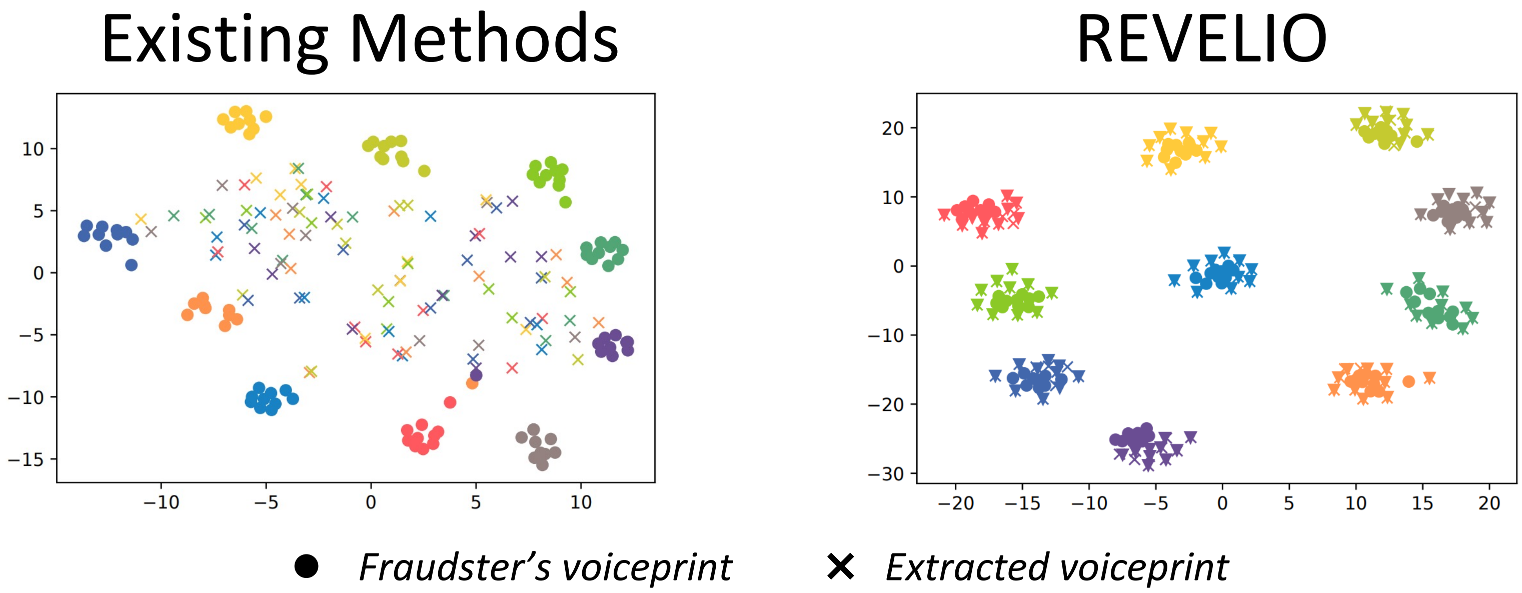


Compared with the Existing Methods



7 times lower

Effectiveness: Identify the Fraudster

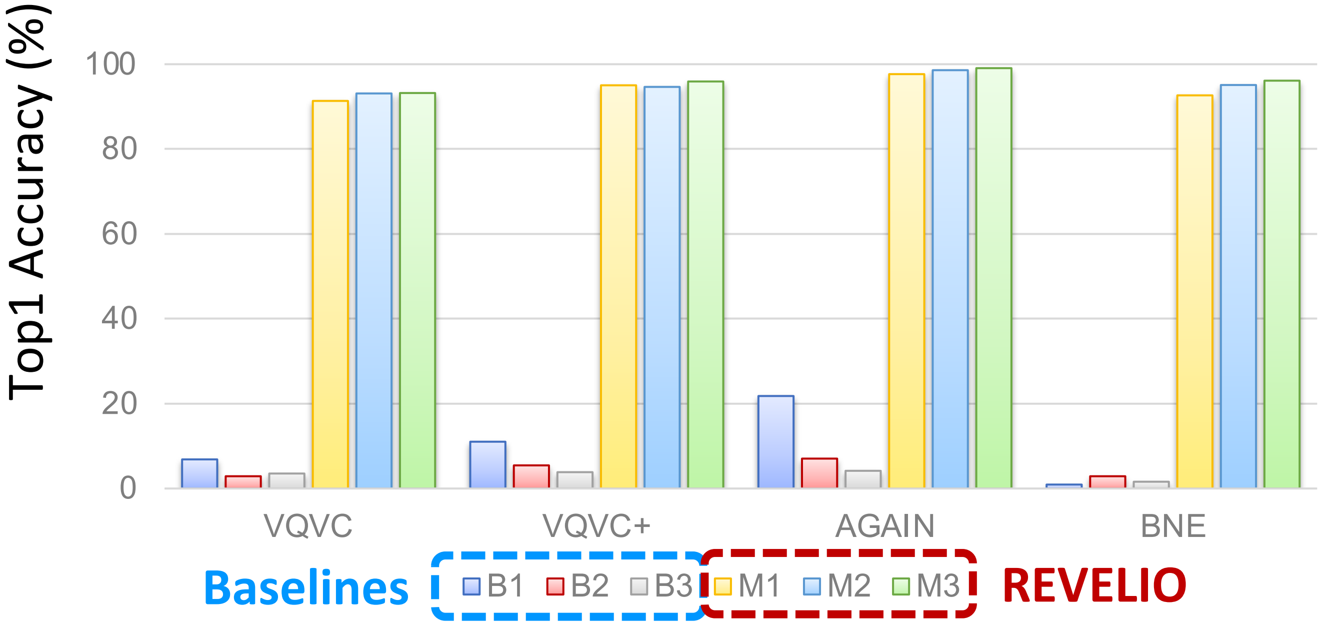

Existing Methods

REVELIO

● *Fraudster's voiceprint* ✕ *Extracted voiceprint*

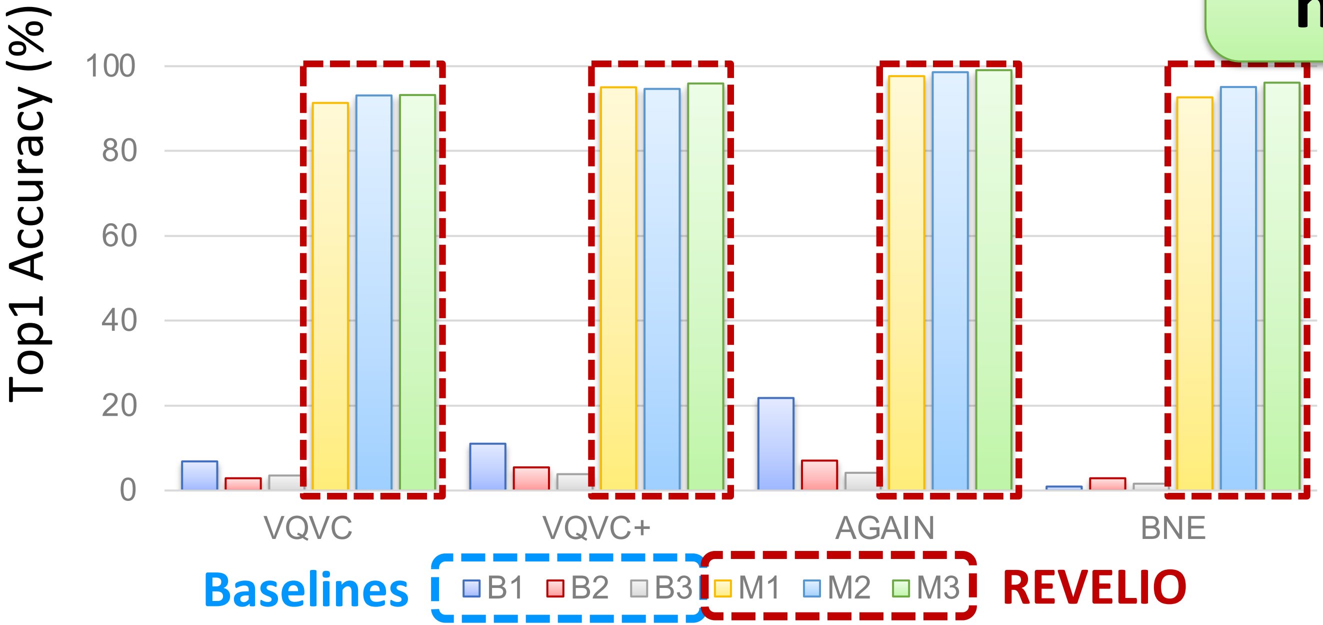
REVELIO can accurately identify the fraudster with the VC audio.

Compared with the Existing Methods

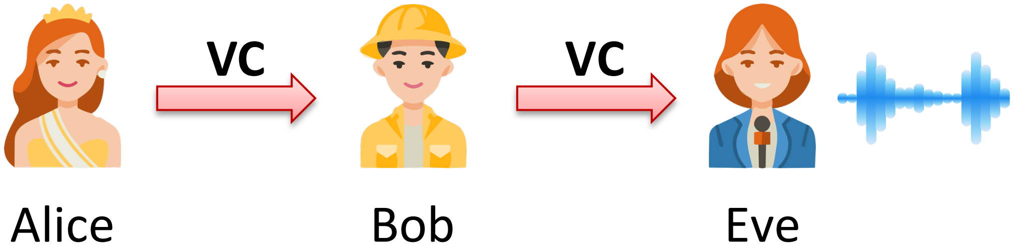


Compared with the Existing Methods

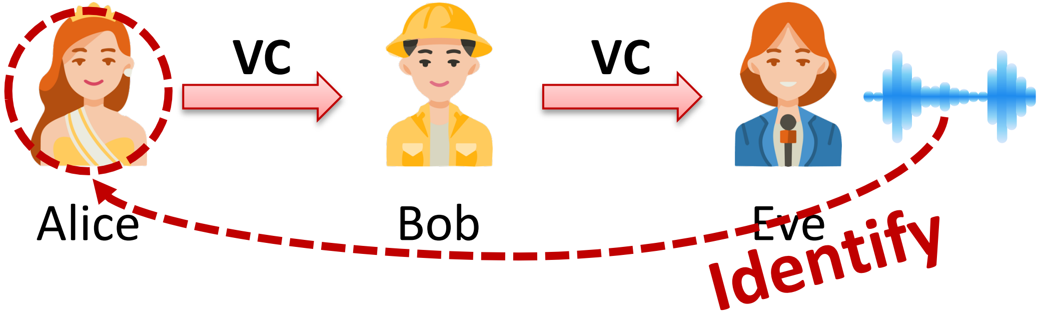
5 times higher



Multiple Voice Conversion

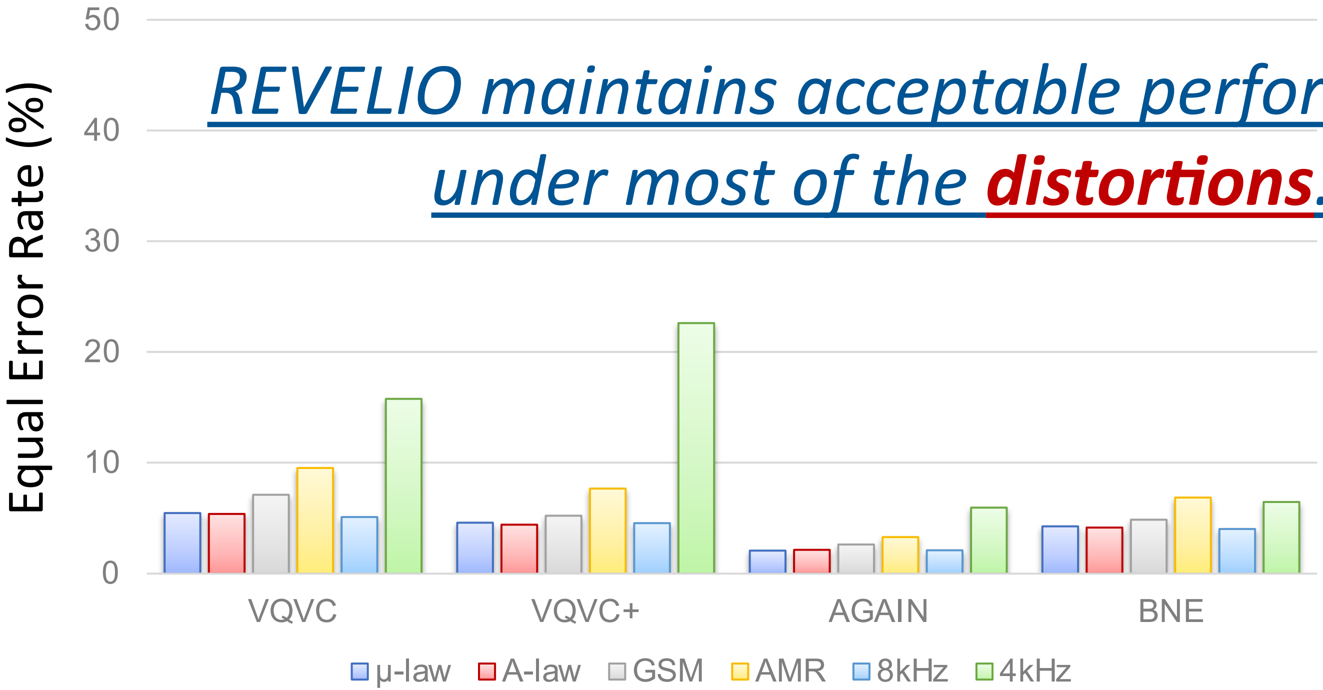


Multiple Voice Conversion



Method	VQVC ×2	VQVC+ ×2	AGAIN ×2	BNE ×2
EER	3.49	3.55	2.27	7.73
Top-1 ACC	96.60	95.64	99.39	70.32
Top-5 ACC	99.65	99.78	100.0	91.67
Top-10 ACC	99.90	99.97	100.0	96.31

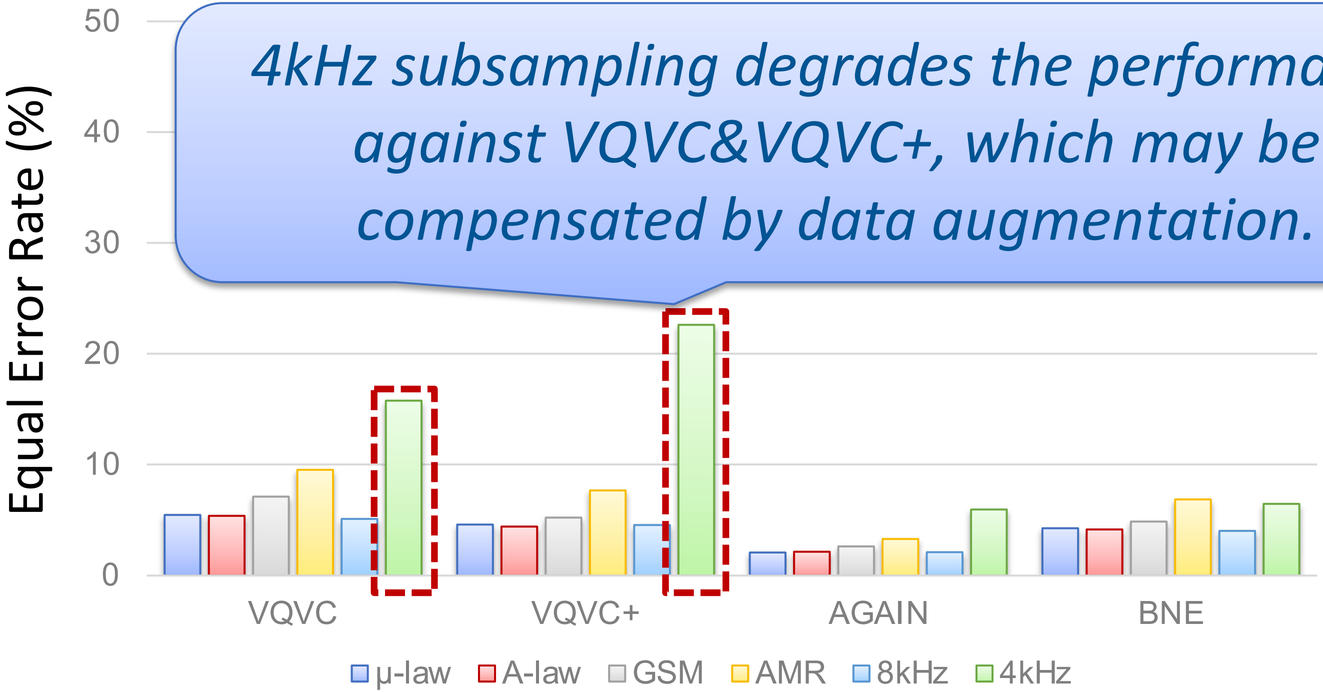
Over-the-telephony Robustness



REVELIO maintains acceptable performances under most of the ***distortions**.*

PSTN and VoIP codecs **8kHz/4kHz subsampling**

Limitation: against 4kHz Subsampling



4kHz subsampling degrades the performance against VQVC&VQVC+, which may be compensated by data augmentation.

Conclusion

- **First** approach to extract fraudster’s voiceprint from VC audios.
- **First** method that extracts the hidden voiceprint of the source speaker despite of the dominant target voiceprint.
- Validated on 4 VCs, 4 languages, 6 distortions, and a multi-VC adaptive adversary.

REVELIO: REvealing Source VoicEprint ConceaLed by VoIce COnversion

Contact us:

chenyanjiao@zju.edu.cn

USSLAB website:

www.usslab.org