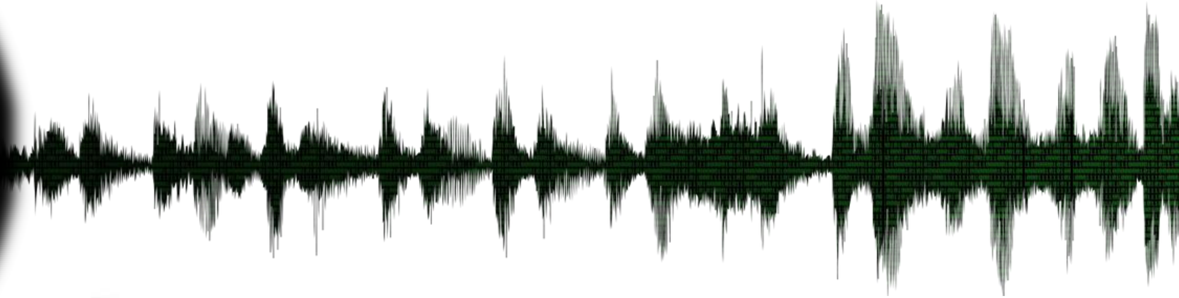
Characterizing the Impact of Audio Deepfakes in the Presence of Cochlear Implants

Magdalena Pasternak, Kevin Warren, Daniel Olszewski, Susan Nittrouer, Patrick Traynor, and Kevin R.B. Butler

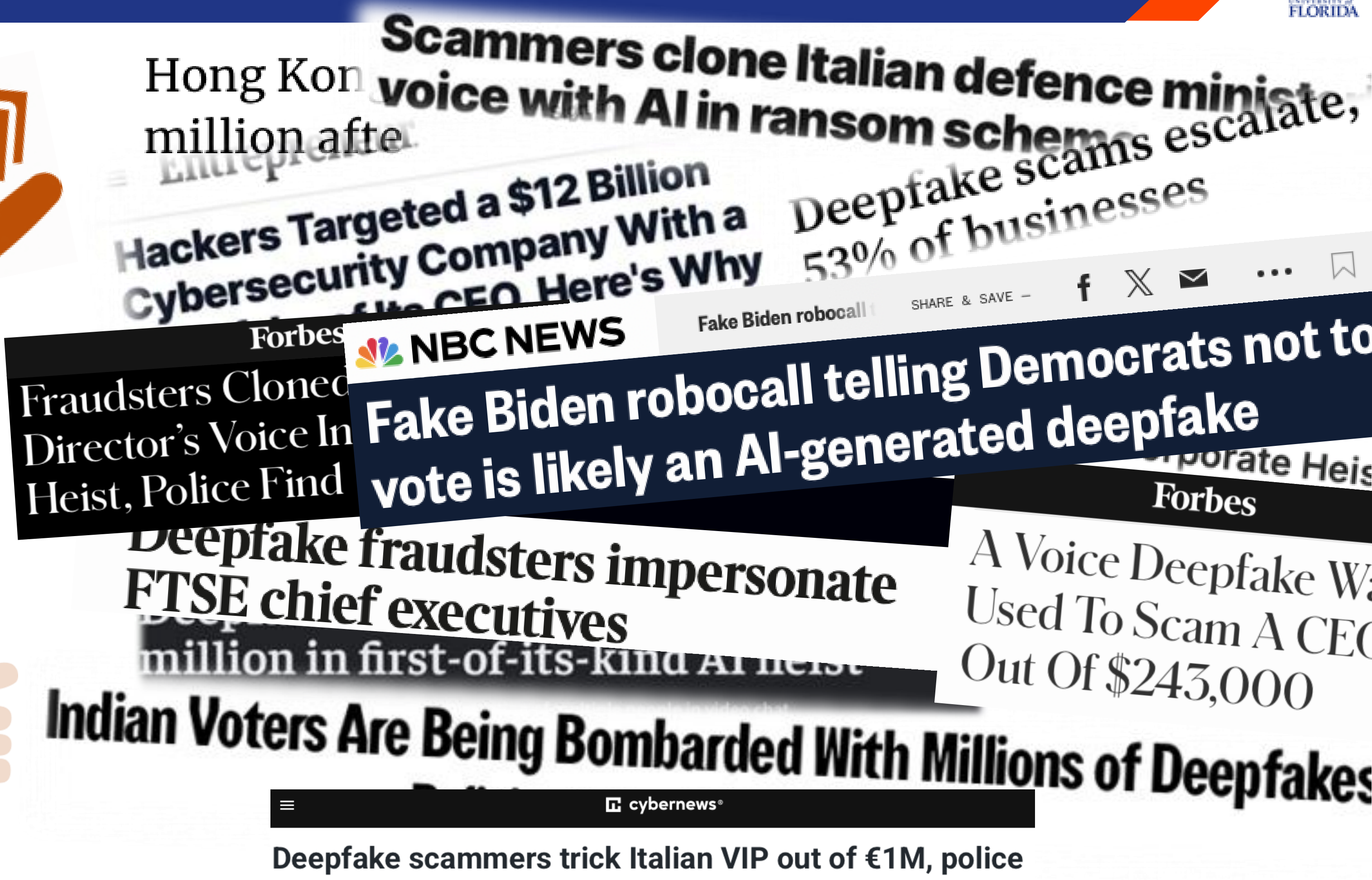
University of Florida




IS IT REAL?

DEEPFAKES


FAKE
NEWS
SHARE


Hong Kon
million afte
Scammers clone Italian defence ministe
voice with AI in ransom scheme
Hackers Targeted a $12 Billion
Cybersecurity Company With a
Here's Why
Deepfake scams escalate,
53% of businesses
Forbes
Fraudsters Clone
Director's Voice In
Heist, Police Find
NBC NEWS
Fake Biden robocall
SHARE & SAVE
Fake Biden robocall telling Democrats not to
vote is likely an AI-generated deepfake
Forbes
A Voice Deepfake W
Used To Scam A CEO
Out Of $243,000
Deepfake fraudsters impersonate
FTSE chief executives
million in first-of-its-kind AI heist
Indian Voters Are Being Bombarded With Millions of Deepfakes
cybernews
Deepfake scammers trick Italian VIP out of €1M, police

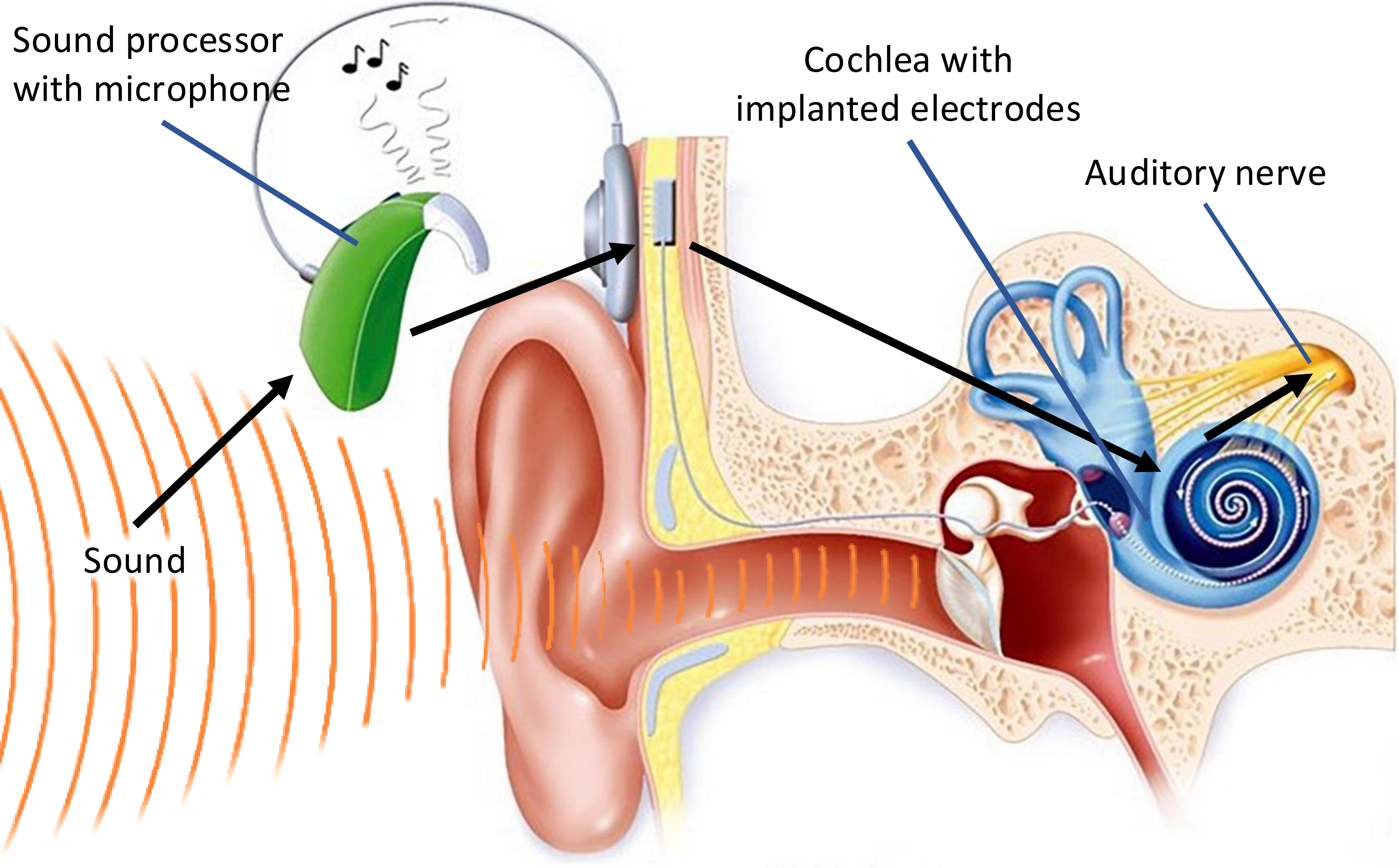
Scammers clone Italian defence ministe
voice with AI in ransom scheme
Deepfake scams escalate,
53% of businesses
12 Billion
ny With a
Here's Why
NEWS
Fake Biden robocall
SHARE & SAVE
ake Biden robocall telling Democrats not to
vote is likely an AI-generated deepfake
Heist, Police
Deepfake fraudsters impersonate
FTSE chief executives
in first-of-its-kind AI heist
Forbes
A Voice Deepfake W
Used To Scam A CE
Out Of $243,000
oters Are Being Bombarded With Millions of Deepfakes
cybernews
Deepfake scammers trick Italian VIP out of €1M, police
FAKE NEWS
SHARE

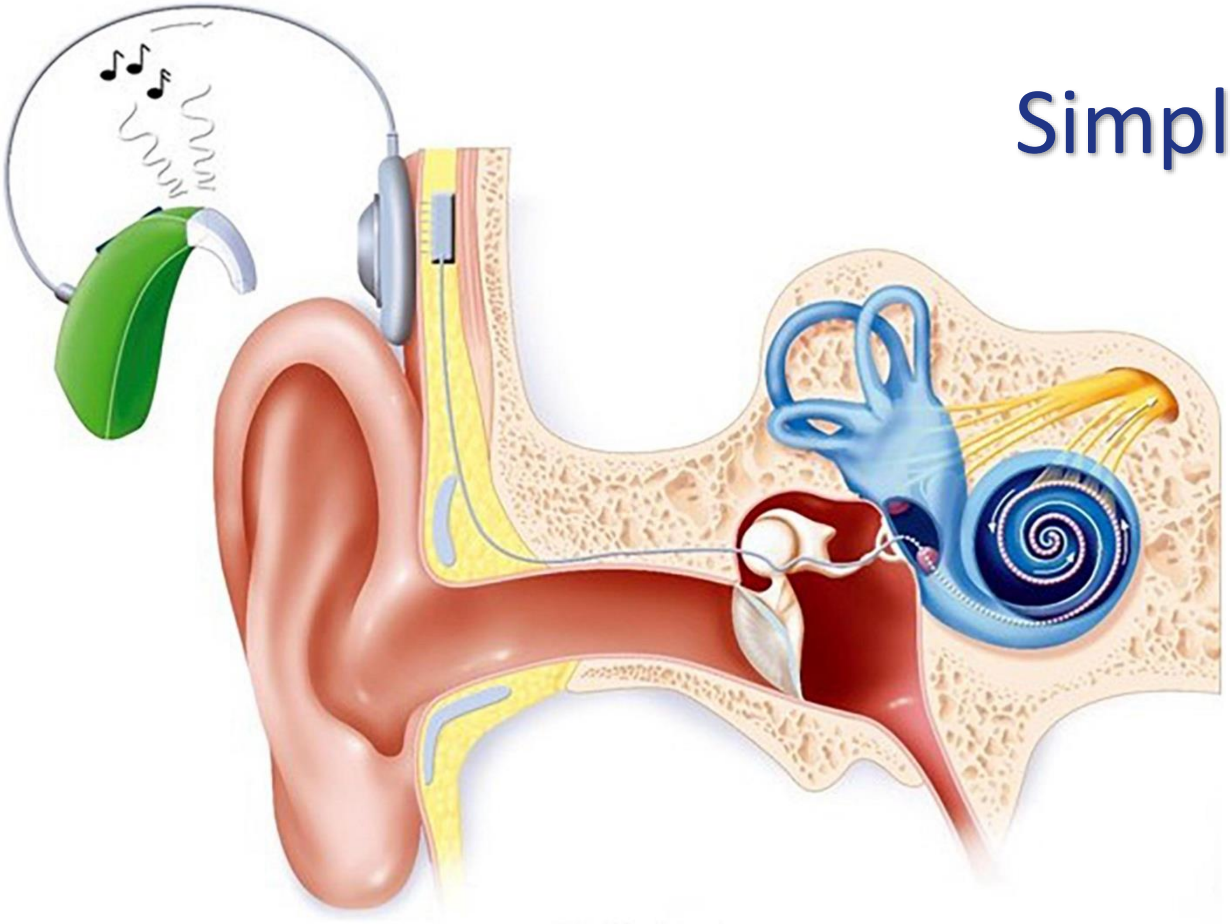
DEEPFAKES


FAKE
NEWS
SHARE


Hong Ko
milli
mmers clone Italian defence ministe
AI in ransom scheme
Deepfake scams escalate,
53% of businesses
Hack
Cyber
Fake Biden robocall
SHARE & SAVE
robocall telling Democrats no
ly an AI-generated deepfake
Fraudster
Director's Vo
Heist, Police Fi
Deepfake fra
FTSE chief ex
million in fi
personate
Forbes
A Voice Deepfake W
Used To Scam A CE
Out Of $243,000
Indian Voters Ar
ing Bo
barded With Millions of Deepfakes
cybernews
De
mmers trick Italian VIP out of €1M, police

Cochlear Implants (CIs)





Simplify & Compress Sound

Difficulties Perceiving:

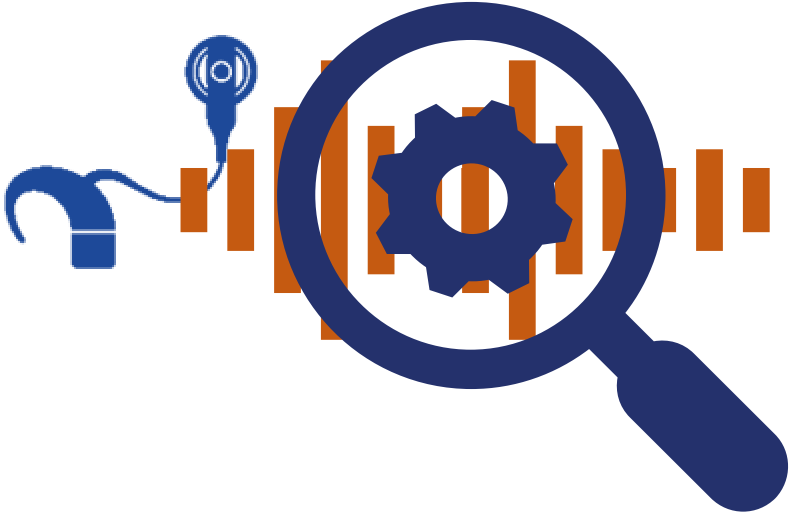
- Pitch
- Tone

Our Research Questions

RQ1. How susceptible are CI users to audio deepfake attacks?

RQ1. How susceptible are CI users to audio deepfake attacks?

RQ1. How susceptible are CI users to audio deepfake attacks?



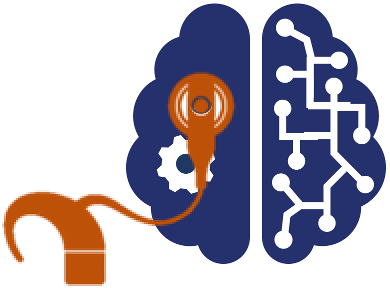
RQ1. How susceptible are CI users to audio deepfake attacks?

RQ2. How effective are automated models on CI-simulated audio?

Our Research Questions

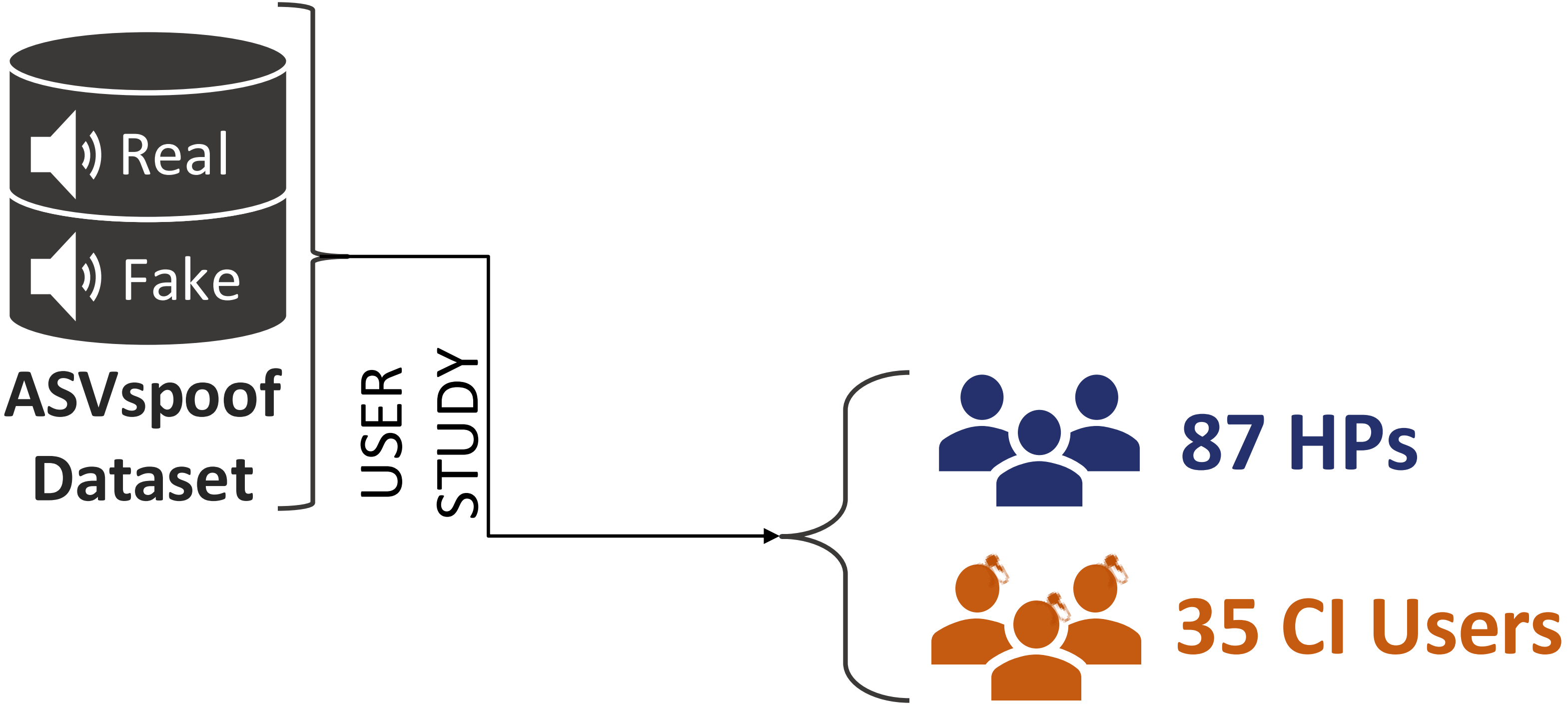
RQ1. How susceptible are CI users to audio deepfake attacks?

RQ2. How effective are automated models on CI-simulated audio?

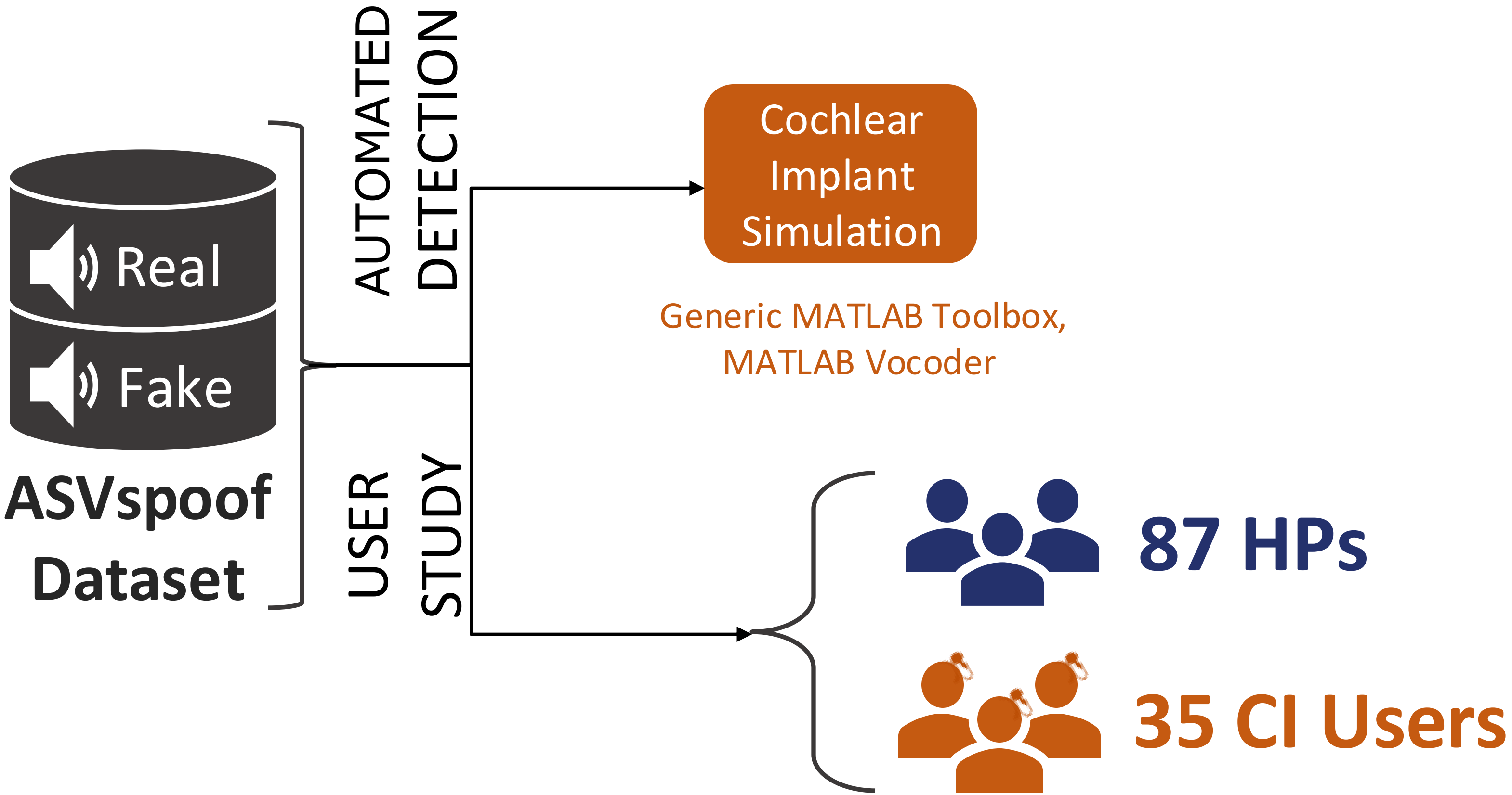


RQ3. Can these models be used as substitutes for CI users?

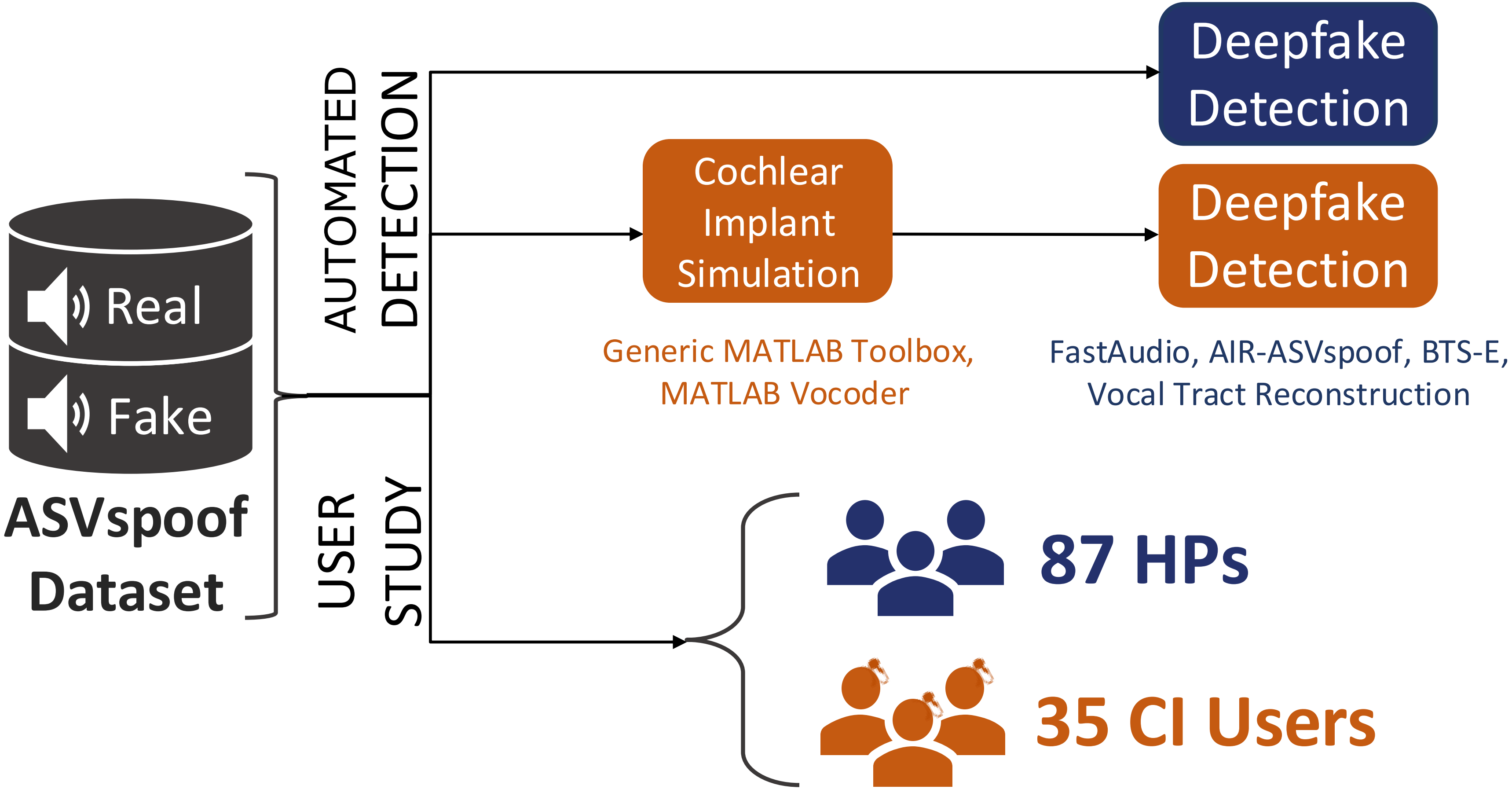
Methodology

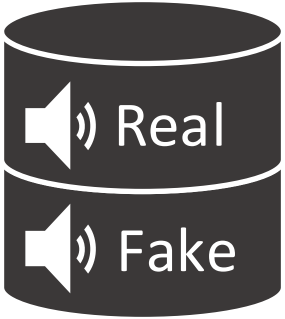


Methodology



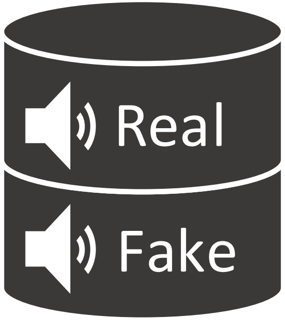
Methodology





ASVspoof Dataset

Automatic **S**peaker **V**erification **spoof**ing detection

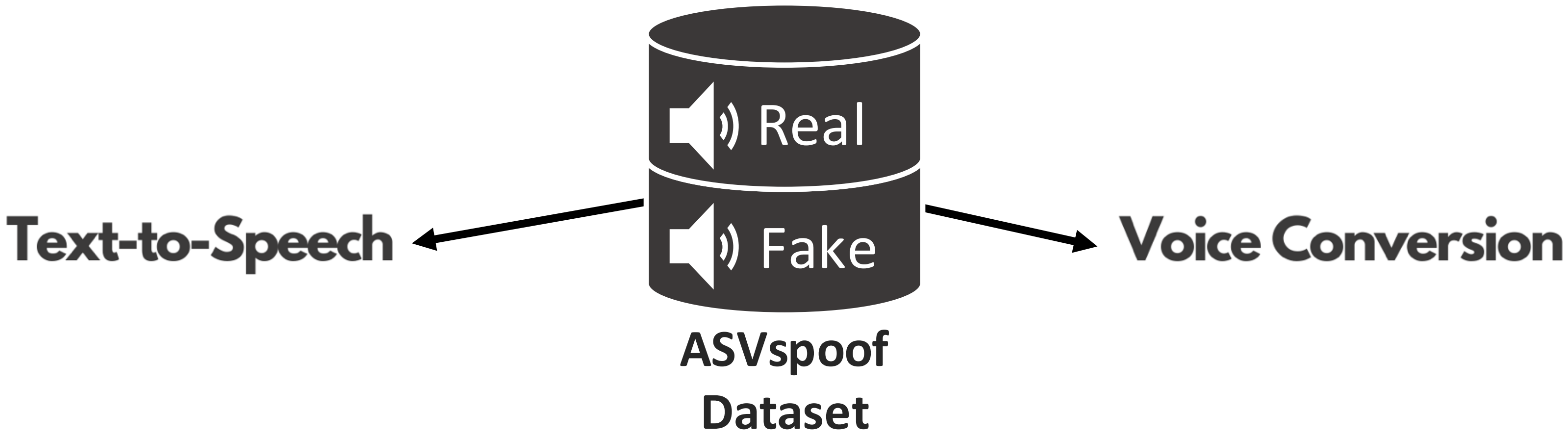


ASVspoof Dataset

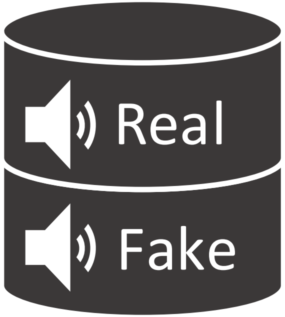
Automatic **S**peaker **V**erification **spoof**ing detection

Speaker Verification

Spoofing Attack Detection


Real
Fake
Text-to-Speech
Voice Conversion
ASVspoof
Dataset

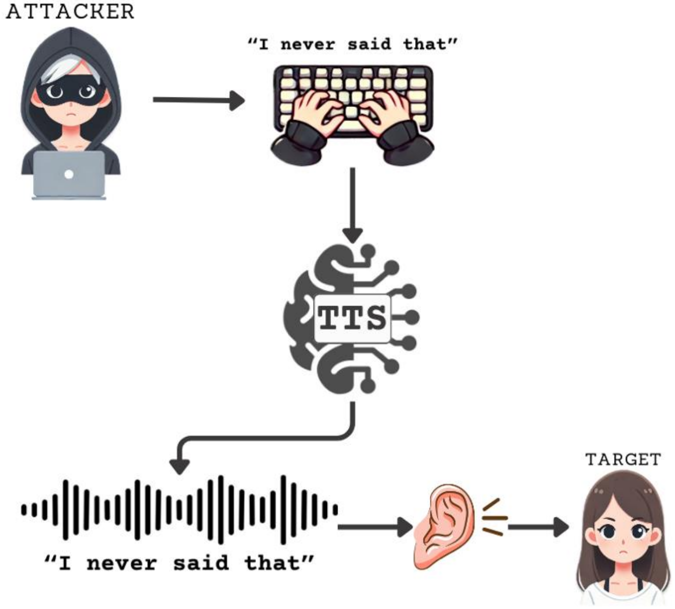
Types of Audio Deepfakes Generation Methods



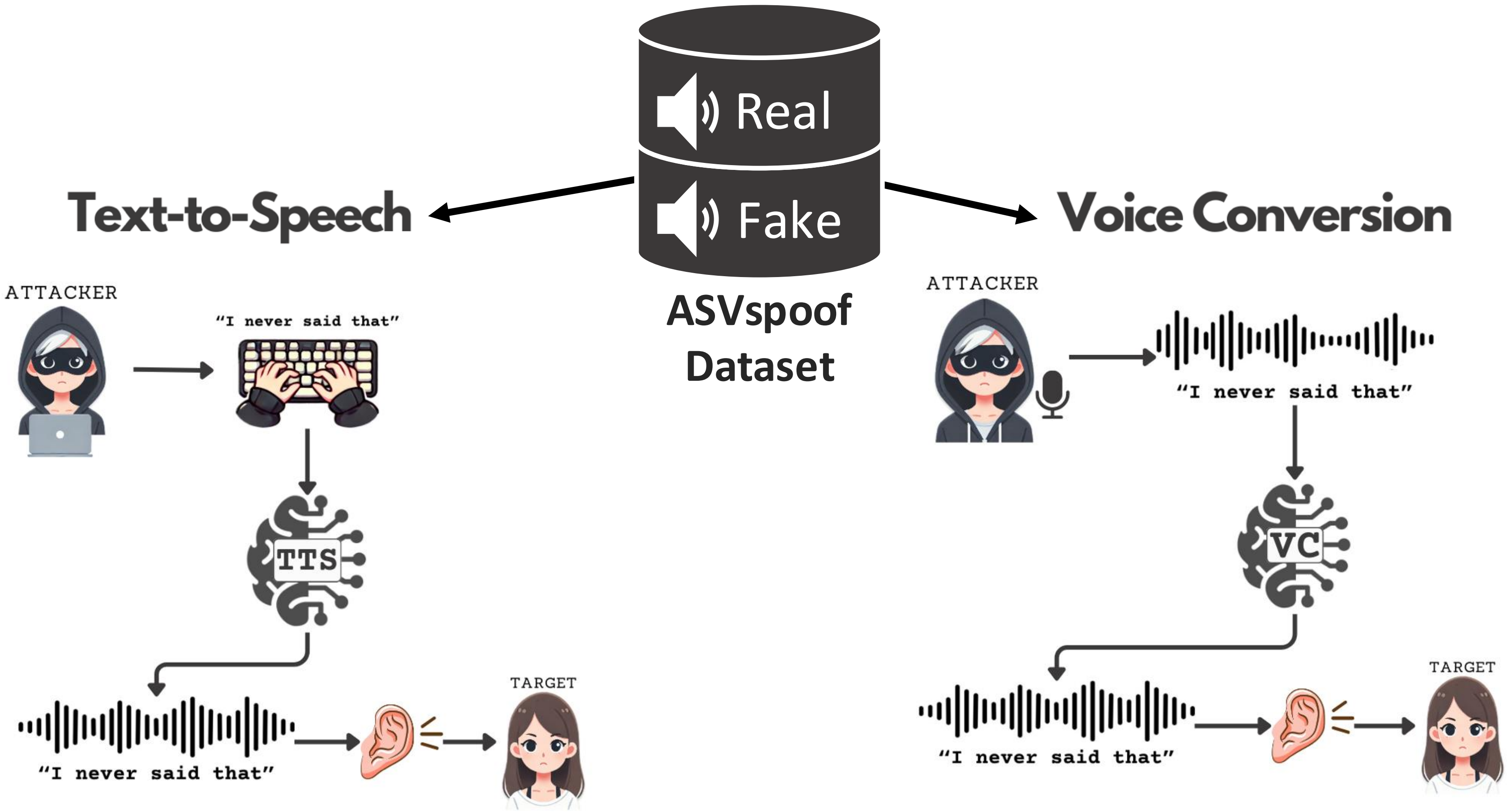
ASVspoof Dataset

Text-to-Speech

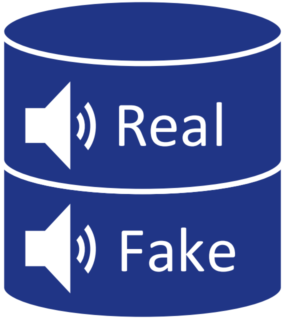
Voice Conversion



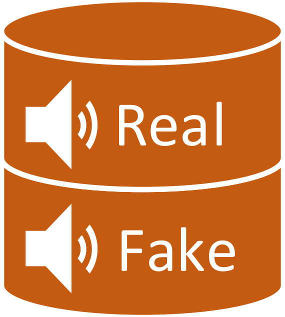
Types of Audio Deepfakes Generation Methods



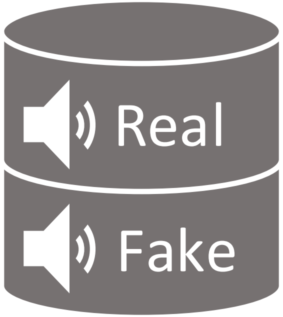
Automated Detection Results – RQ2



Original Dataset

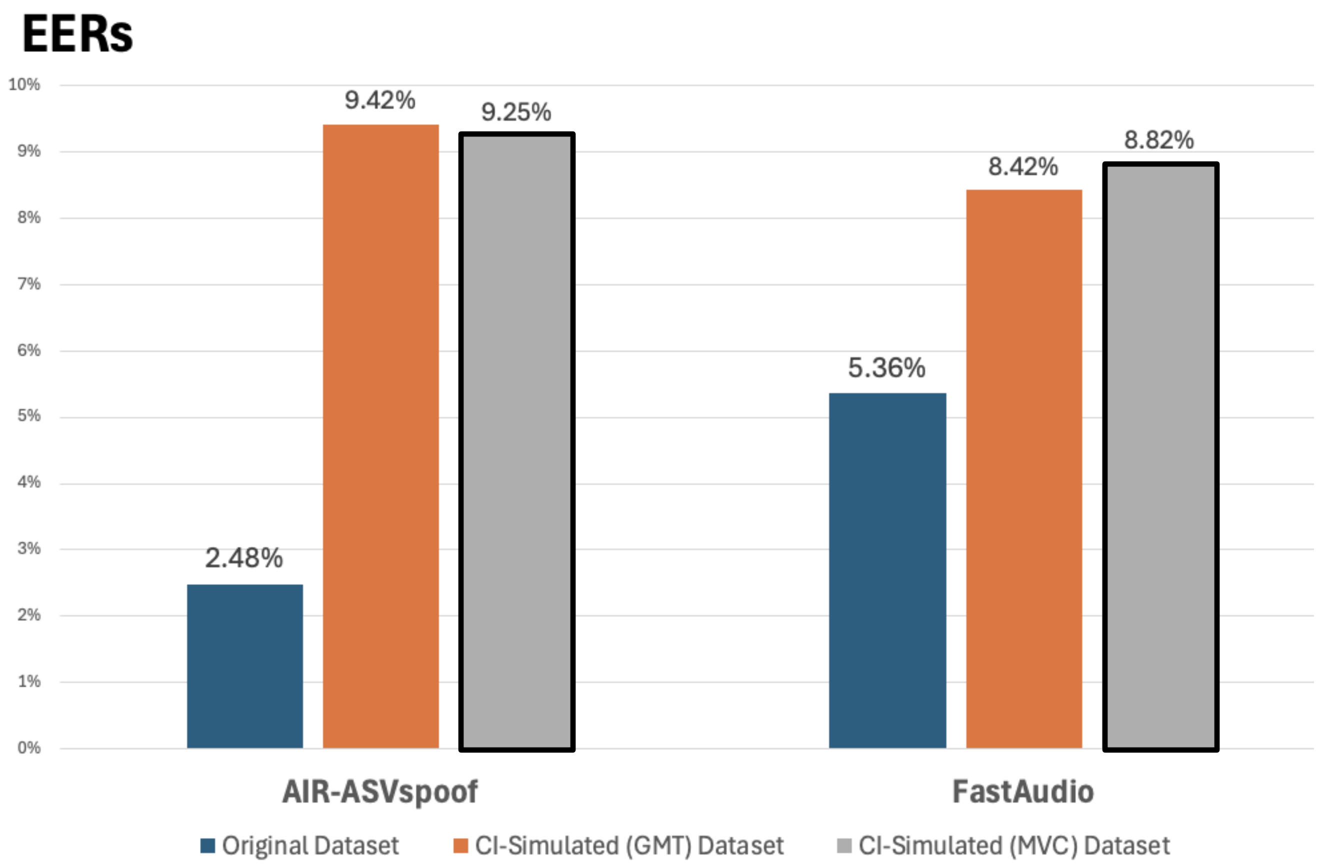


CI-simulated (Generic MATLAB Toolbox) Dataset



CI-simulated (MATLAB Vocoder) Dataset

Automated Detection Results – RQ2



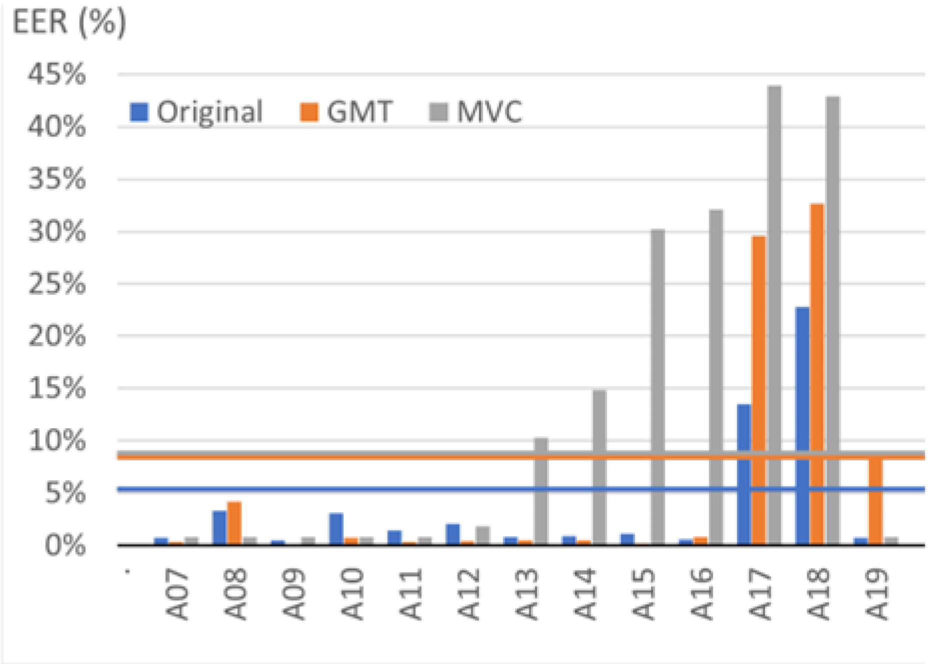
Automated Detection Results – RQ2

Spoofing attack systems and approaches in the ASVspoof2019 dataset

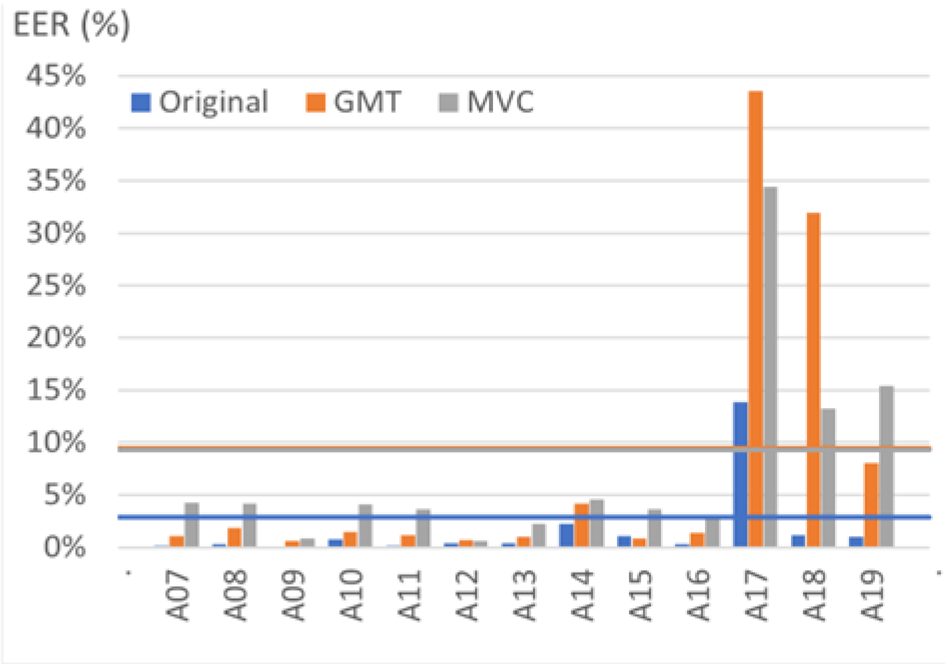
Attack	System	Approach
A01	TTS	neural waveform model
A02	TTS	vocoder
A03	TTS	vocoder
A04	TTS	waveform concatenation
A05	VC	vocoder
A06	VC	spectral filtering
A07	TTS	vocoder+GAN
A08	TTS	neural waveform
A09	TTS	vocoder
A10	TTS	neural waveform
A11	TTS	griffin lim
A12	TTS	neural waveform
A13	TTS&VC	waveform conc. & filt.
A14	TTS&VC	vocoder
A15	TTS&VC	neural waveform
A16	TTS	waveform concatenation
A17	VC	waveform filtering
A18	VC	vocoder
A19	VC	spectral filtering

Automated Detection Results – RQ2

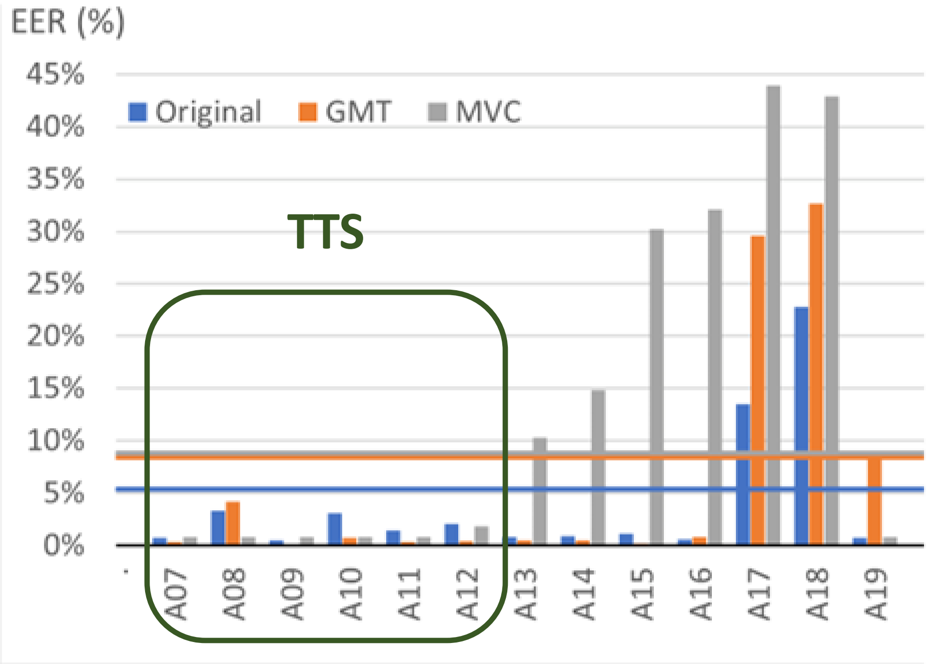
FastAudio



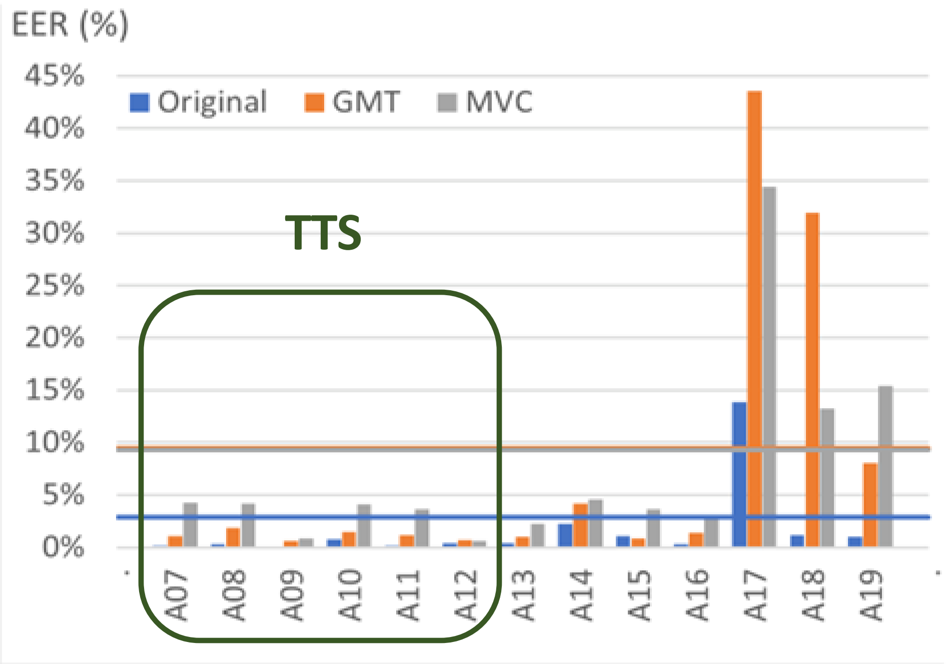
AIR-ASVspoof



FastAudio

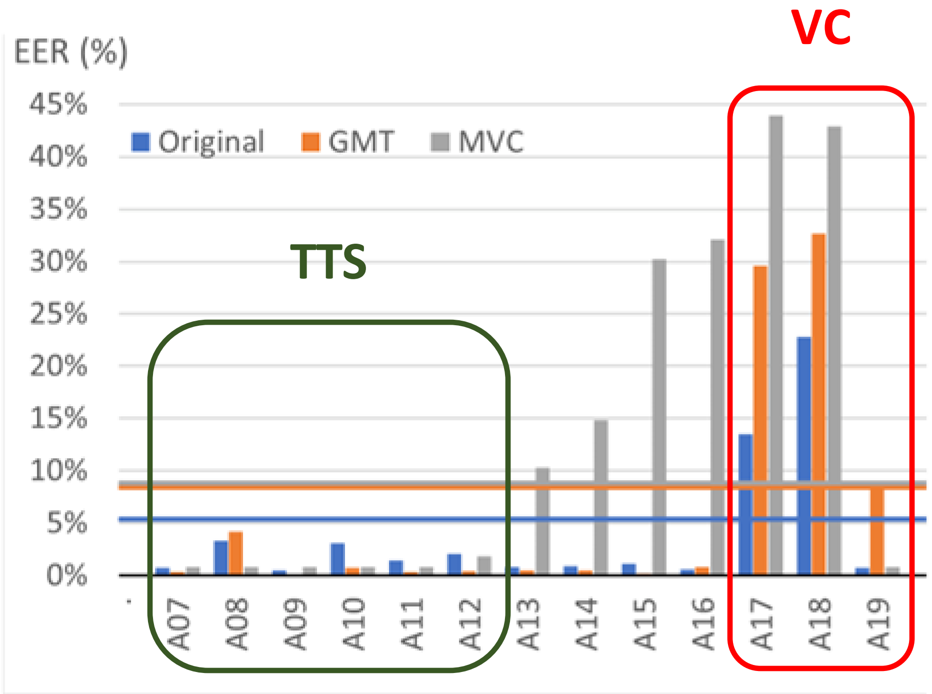


AIR-ASVspoof

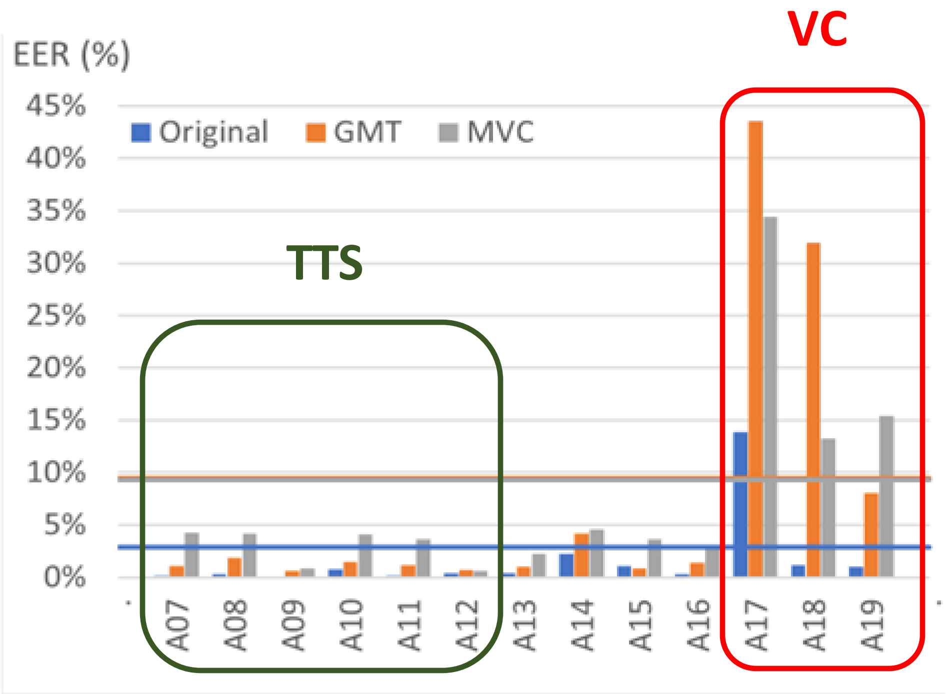


Automated Detection Results – RQ2

FastAudio

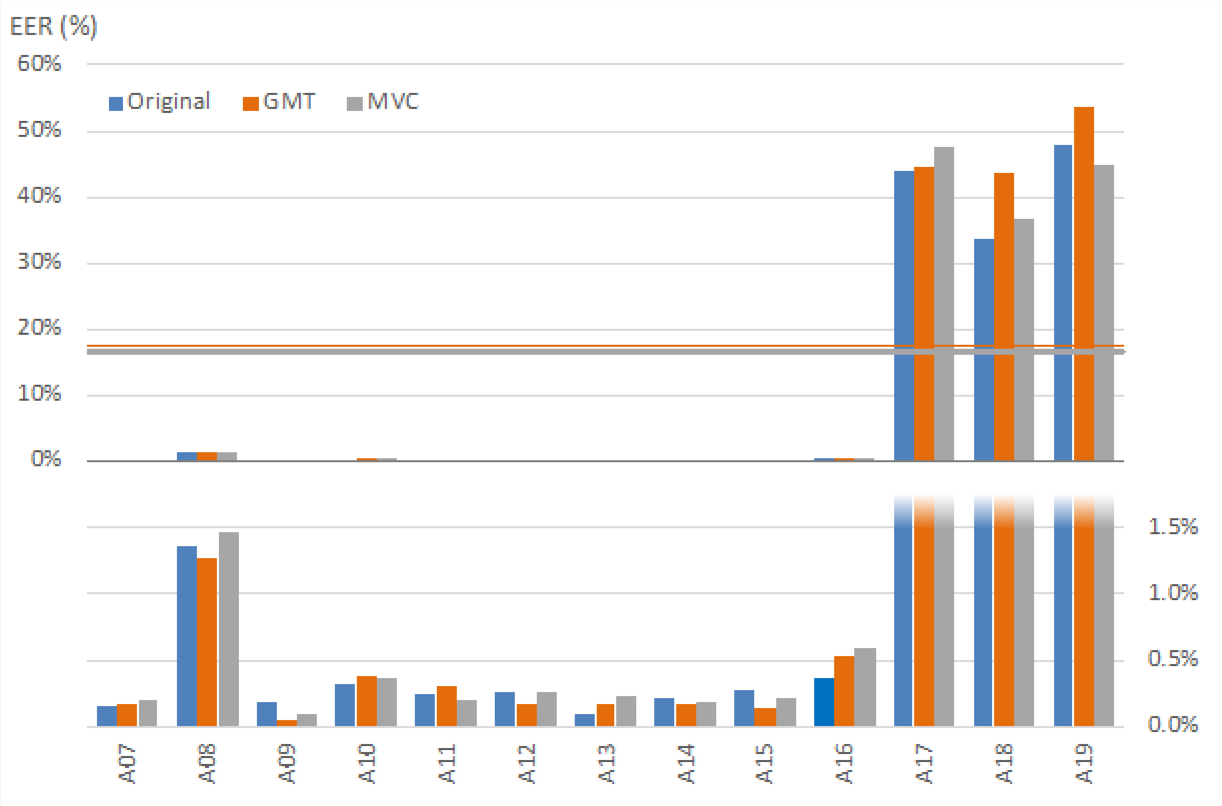


AIR-ASVspoof

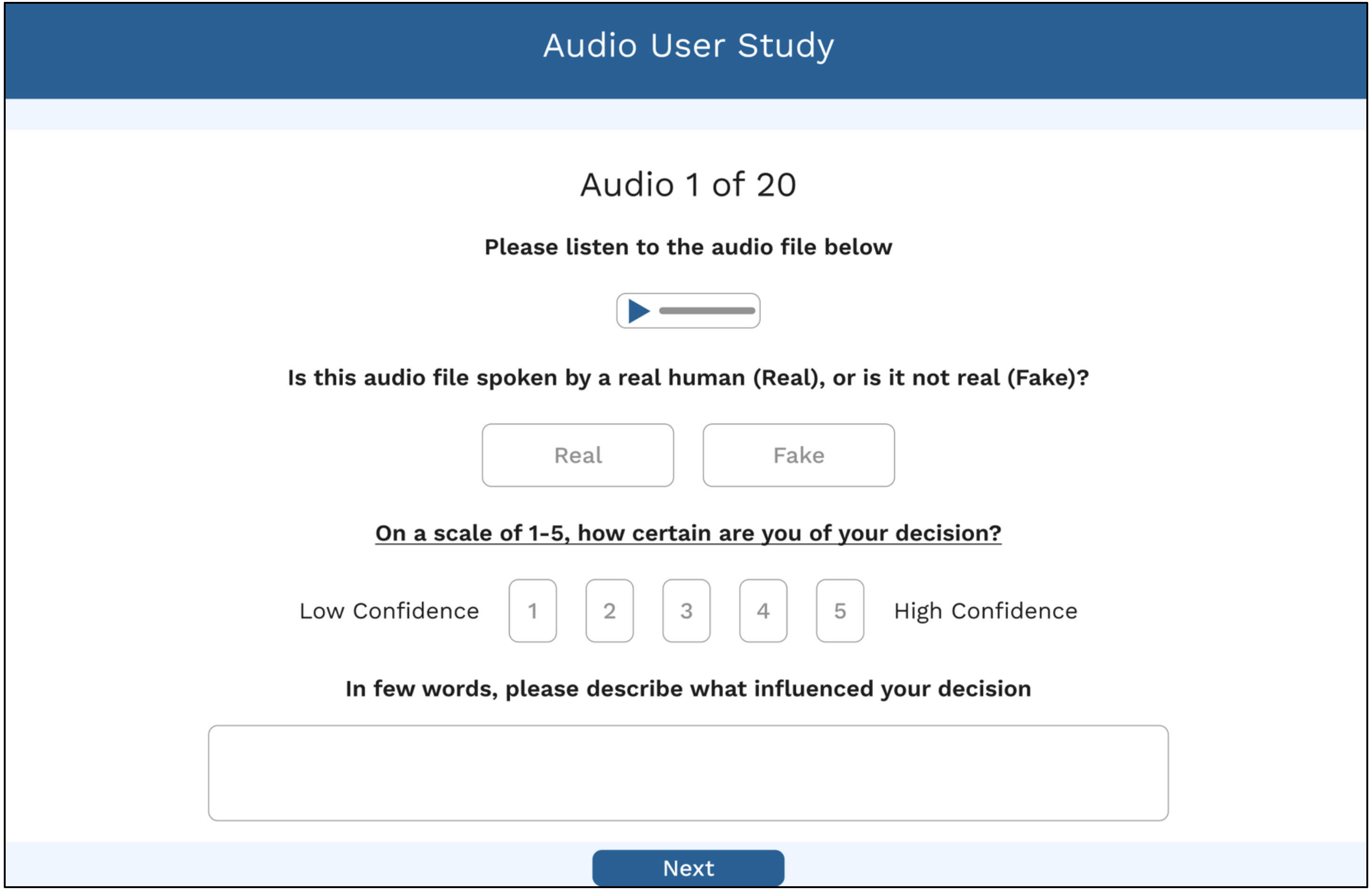


Automated Detection Results – RQ2

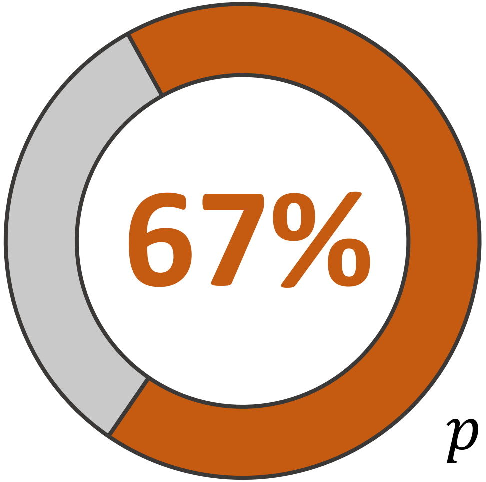
BTS-E



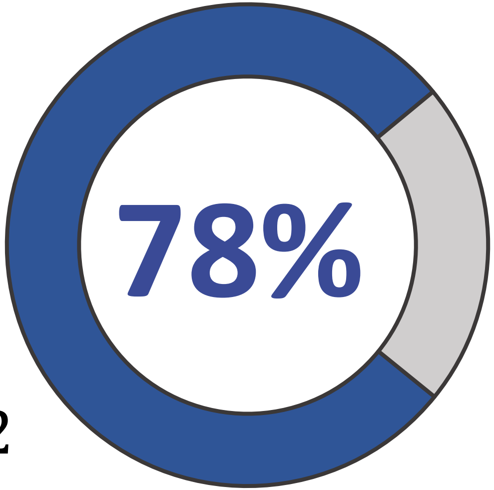
User Study – RQ1



Detection Accuracy



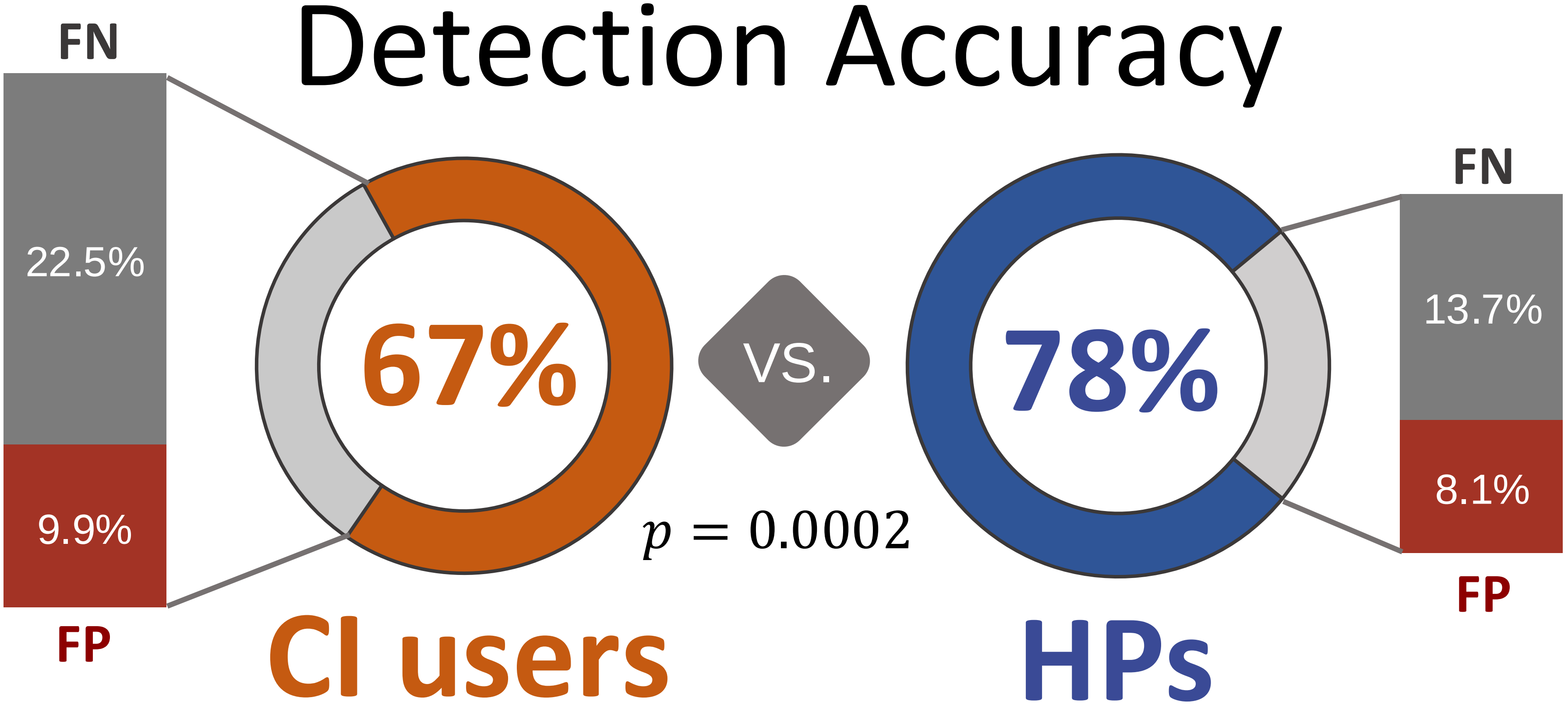
vs.



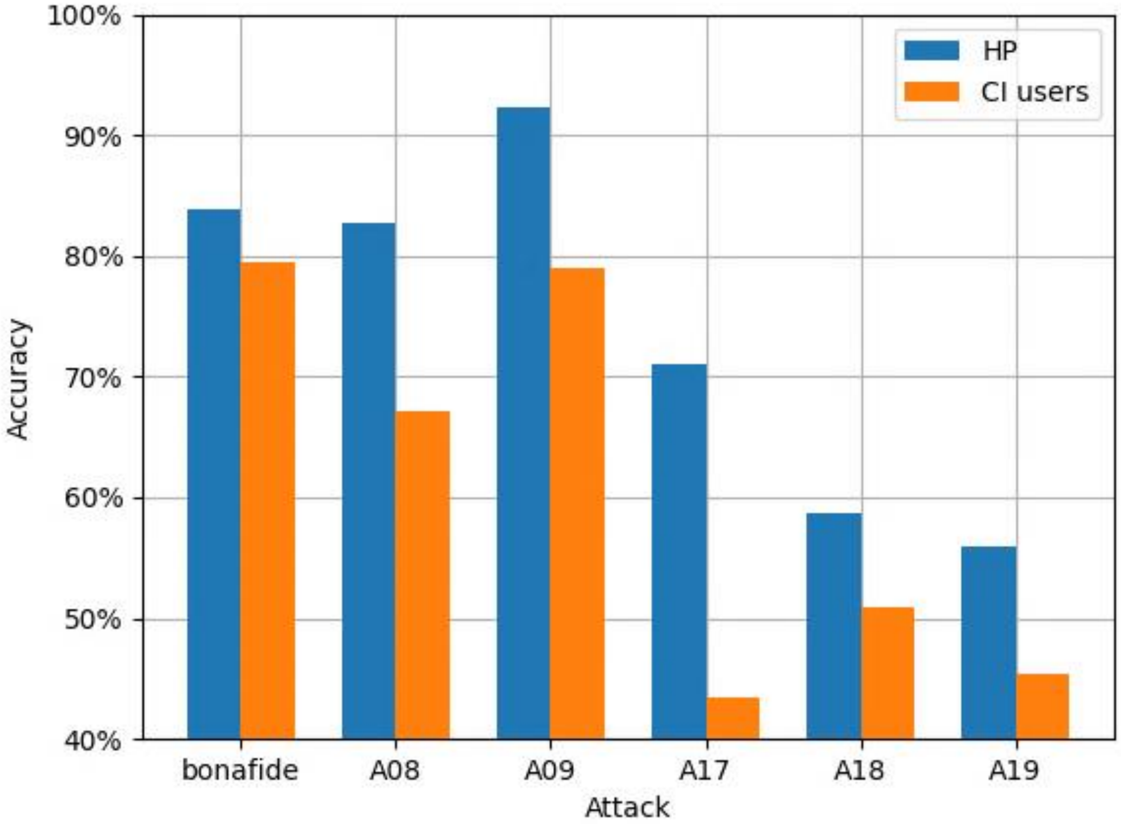
$p = 0.0002$

CI users

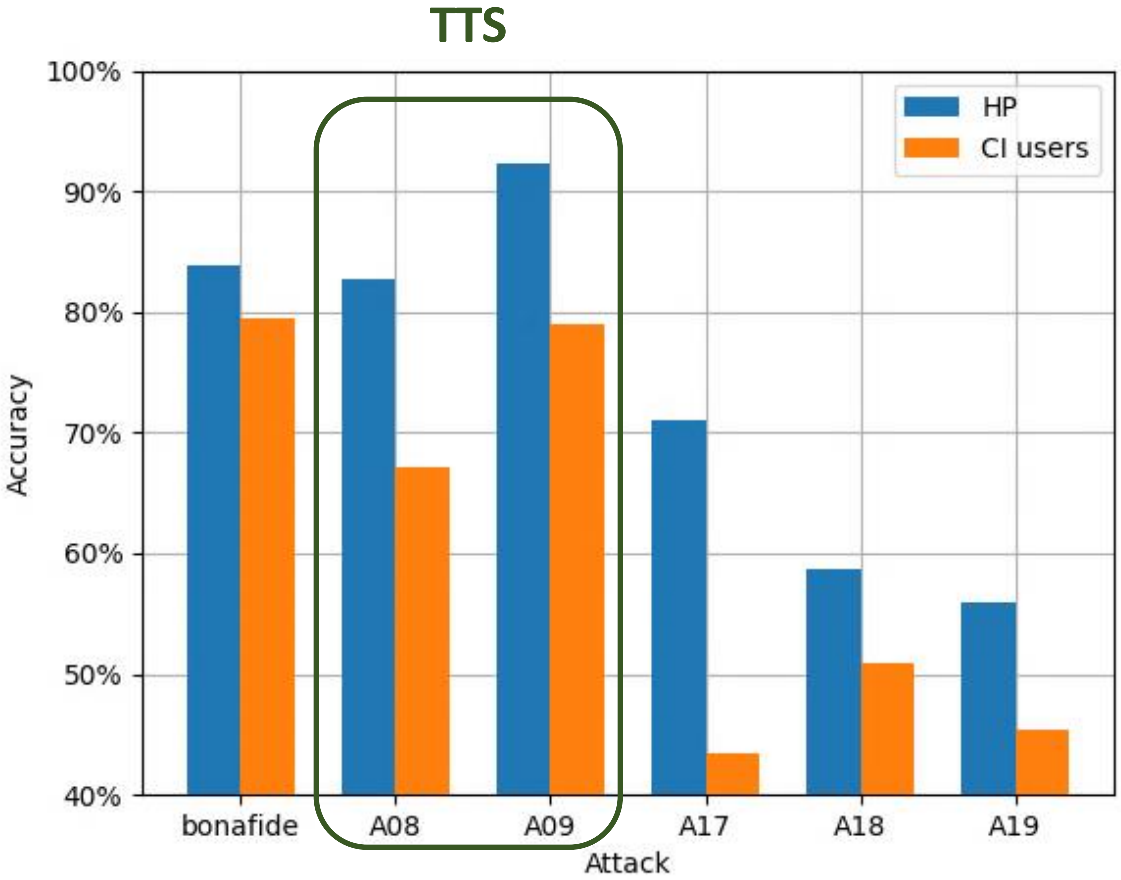
HPs


Detection Accuracy
FN
22.5%
9.9%
FP
67%
CI users
VS.
$p = 0.0002$
78%
HPs
FN
13.7%
8.1%
FP

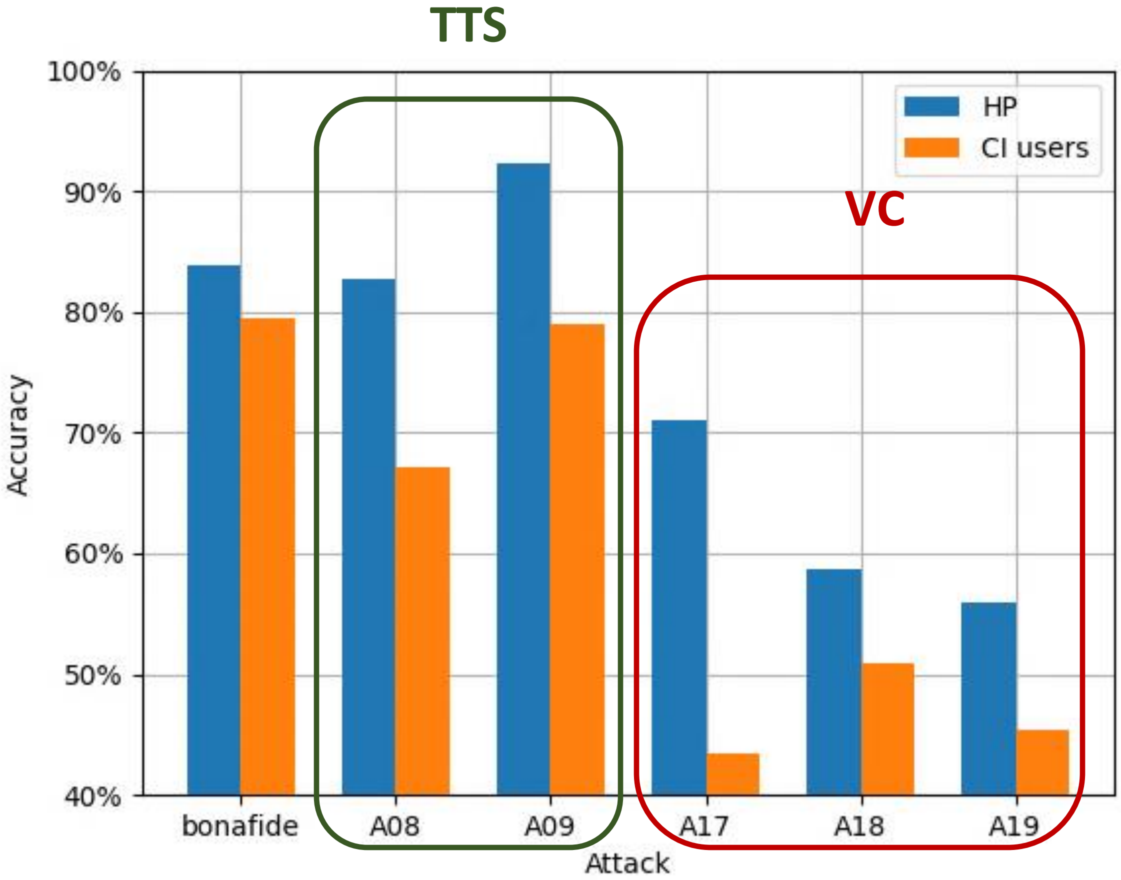
User Study Results – RQ1



User Study Results – RQ1



User Study Results – RQ1

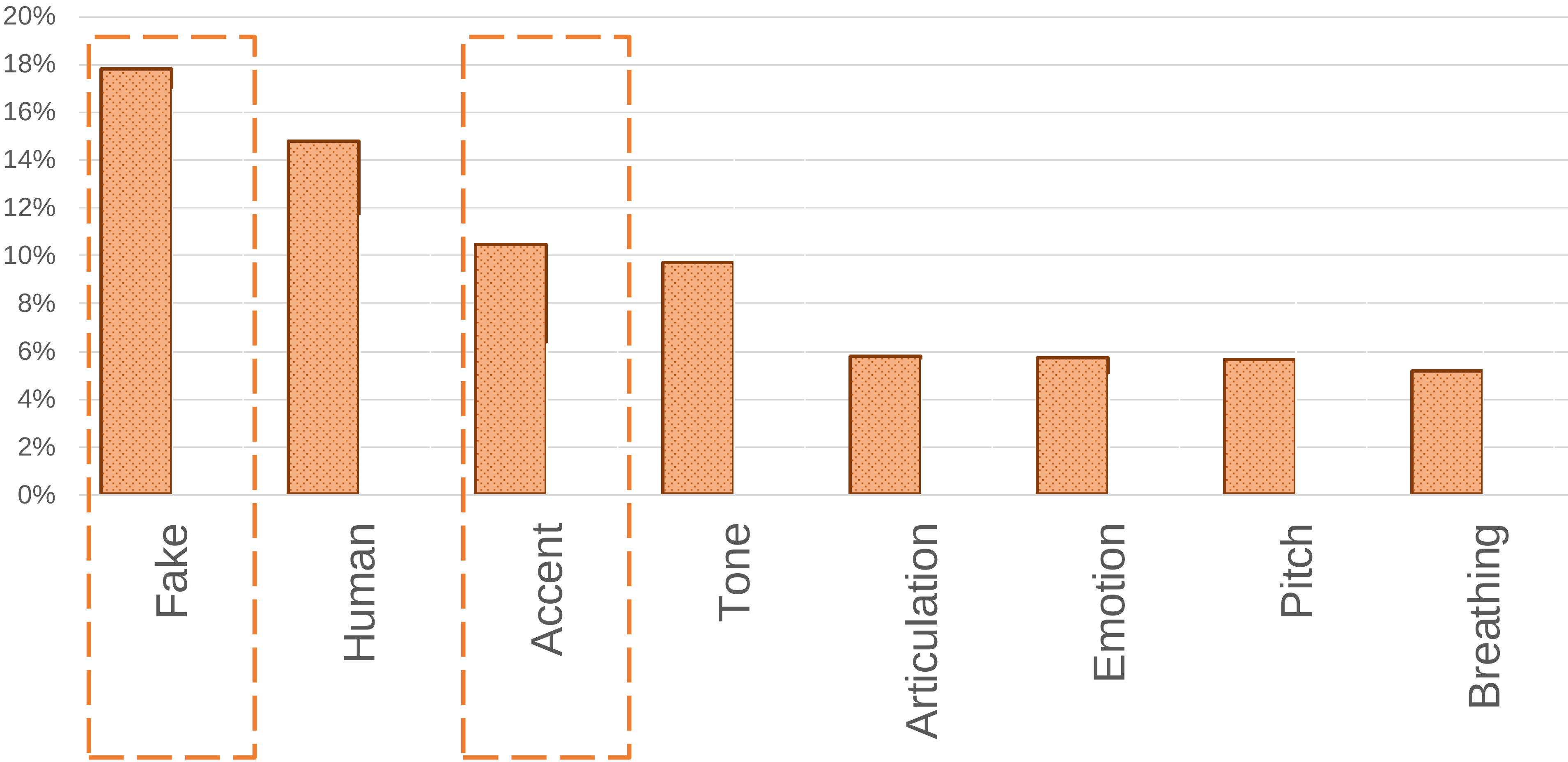


WHY?

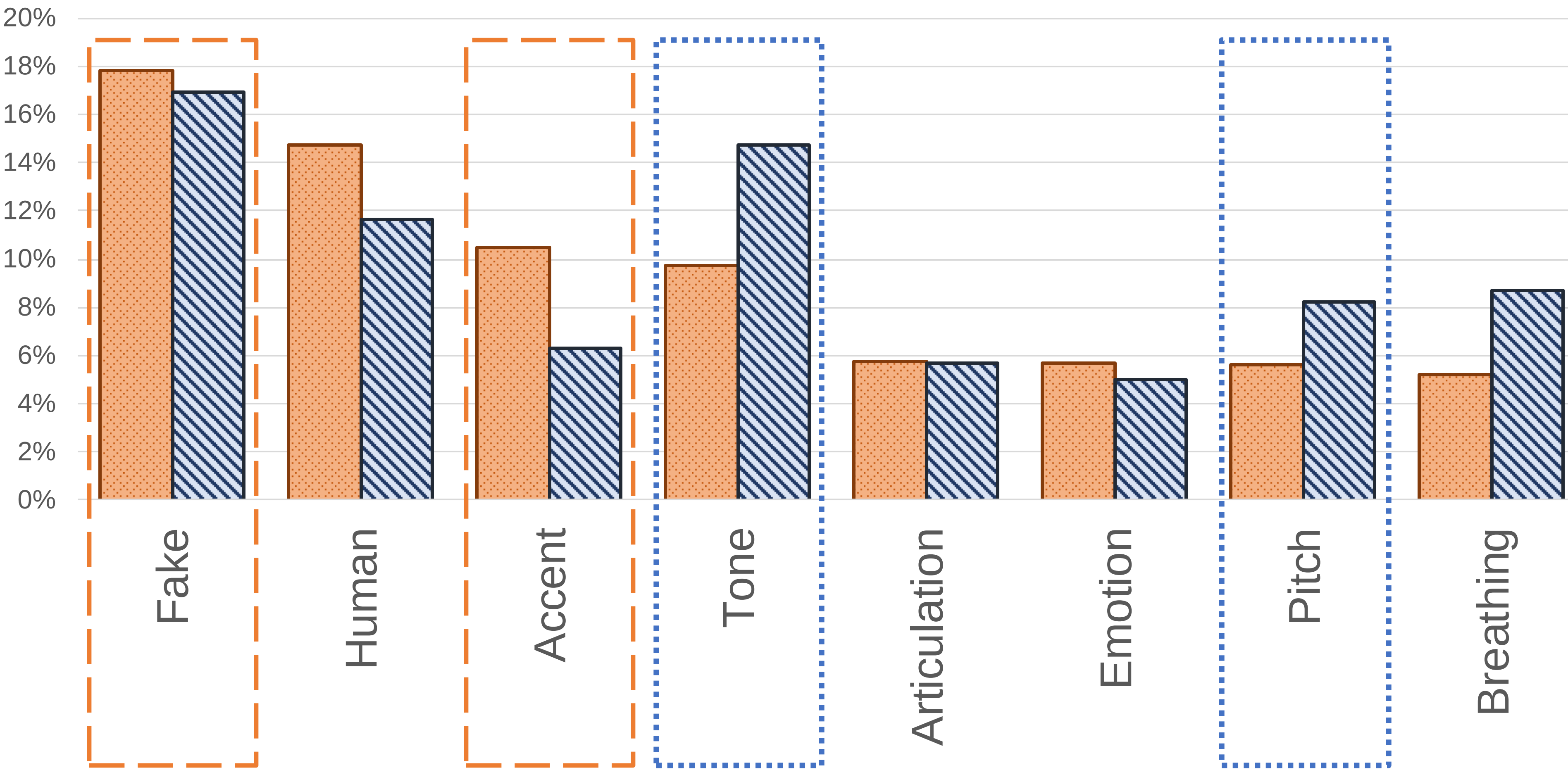
WHY?

Fake
Human
Accent
Tone
Articulation
Emotion
Pitch
Breathing
Guess
Rhytm
Recording
Familiar
Speed
Pauses
Background
Mouth

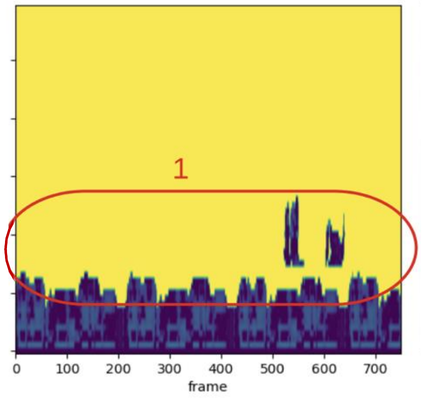
Cue Reliance – RQ1



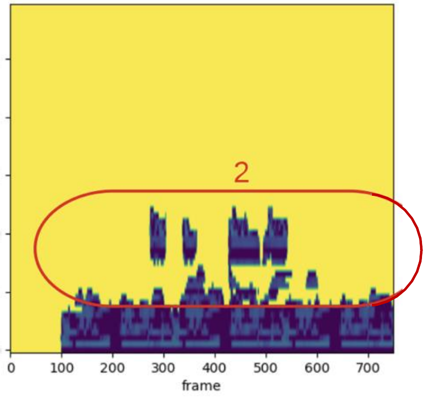
Cue Reliance – RQ1



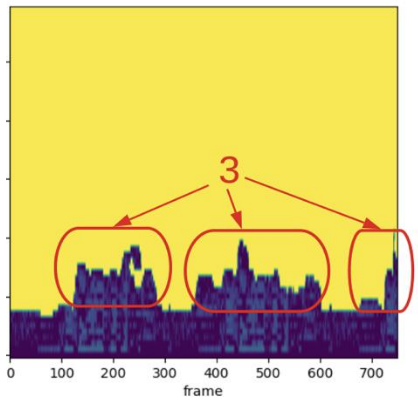
Human vs. Model Evaluation – RQ3



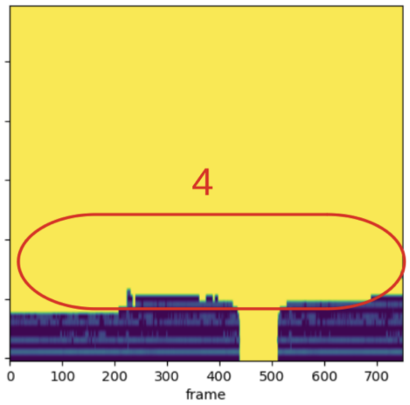
(a) A09 - TTS without CI simulation



(b) A09 - TTS with CI simulation



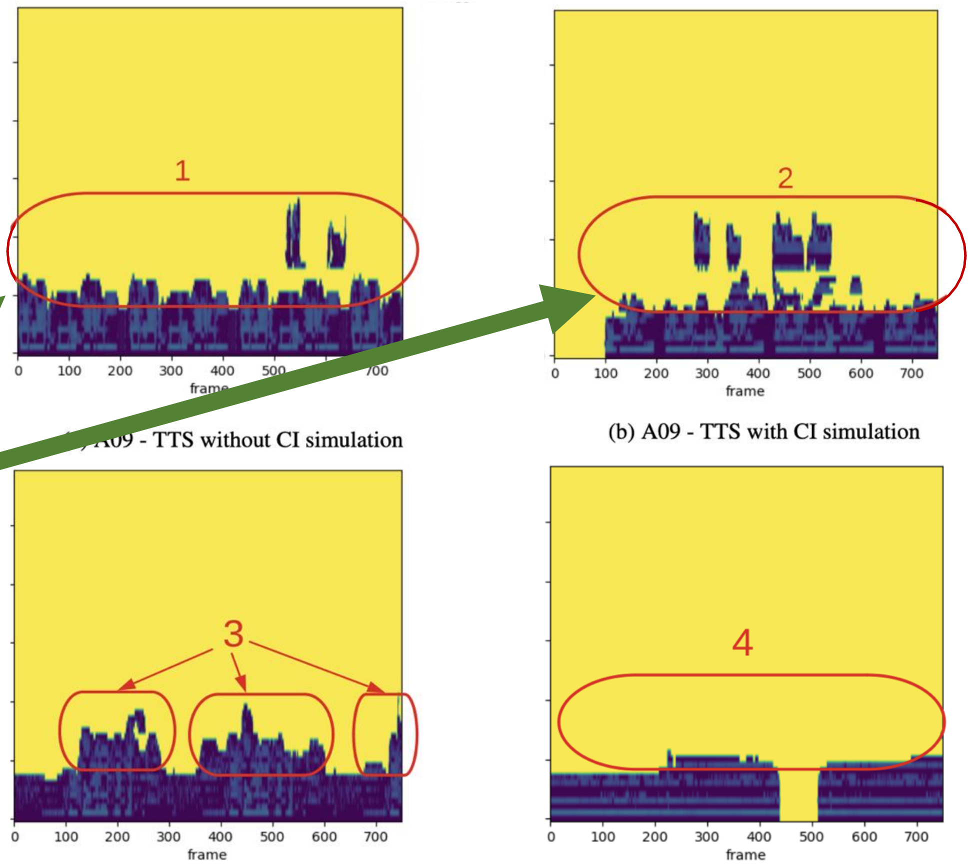
(c) A17 - VC without CI simulation



(d) A17 -VC with CI simulation

Human vs. Model Evaluation – RQ3

TTS



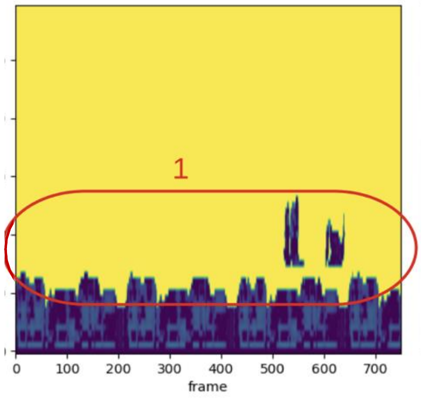
(a) A09 - TTS without CI simulation

(b) A09 - TTS with CI simulation

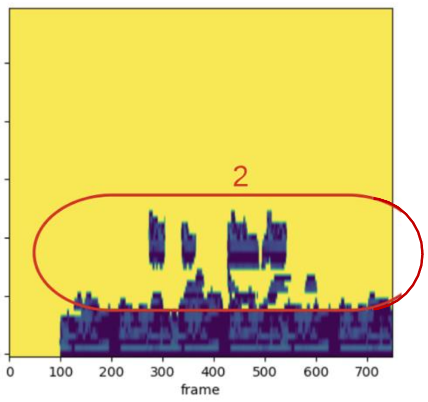
(c) A17 - VC without CI simulation

(d) A17 -VC with CI simulation

Human vs. Model Evaluation – RQ3

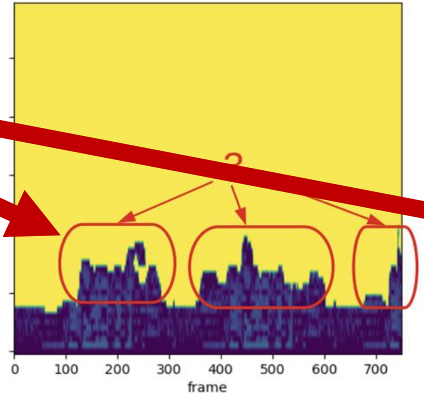


(a) A09 - TTS without CI simulation

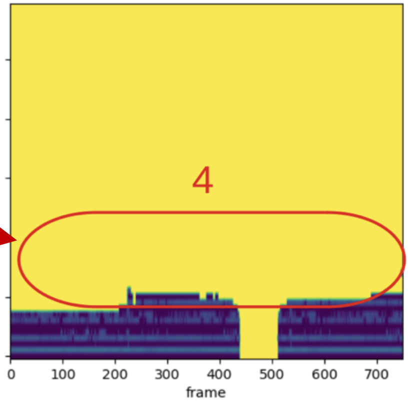


(b) A09 - TTS with CI simulation

VC



(c) A17 - VC without CI simulation



(d) A17 -VC with CI simulation

Conclusions

CI users are disproportionally vulnerable to VC deepfakes

Conclusions

CI users are disproportionally vulnerable to VC deepfakes

Improving the proxy will allow the enhancement of assistive deepfake detectors

Conclusions

CI users are disproportionally vulnerable to VC deepfakes

Improving the proxy will allow the enhancement of assistive deepfake detectors

Awareness & Education Programs

Conclusions

CI users are disproportionally vulnerable to VC deepfakes

Improving the proxy will allow the enhancement of assistive deepfake detectors

Awareness & Education Programs

Need for effective real-time assistive deepfake detection tools

Thank You

Magdalena Pasternak

mpasternak@ufl.edu