SAS: A speaker verification spoofing database containing diverse attacks

Zhizheng Wu[1], Ali Khodabakhsh[2], Cenk Demiroglu[2], Junichi Yamagishi[1,3], Daisuke Saito[4], Tomoki Toda[5], Simon King[1]

[1]*University of Edinburgh, United Kingdom*

[2]*Ozyegin University, Turkey*

[3]*National Institute of Informatics, Japan*

[4]*University of Tokyo, Japan*

[5]*Nara Institute of Science and Technology, Japan*

Acknowledgement: Thank Prof. Zhen-Hua Ling, Dr. Ling-Hui Chen, Ms. Li-Juan Liu from University of Science and Technology of China for providing VC-LSP materials

Applications of speaker verification/voice biometric

Speak to Unlock

BARCLAYS

Who we help | What we offer | Insight and Research | Support and information

International Banking / Insight and Research / Manage Your Money / Banking on the power of speech

Banking on the power of speech

Forget passwords and PINs, memorable dates and card verification numbers. Advances in voice biometric technology have made accessing bank accounts through client service centres as simple as small talk.

FURTH

If you wo
product
+44 (

Voice Password

Introducing voice password authentication to the My Vodacom smartphone app for Android and iOS users

- **Your unique, natural voice pattern is your key**
- **Dialect, speaking styles, and pitch are unique to each person**
- **Each voice is distinct. Attempts to impersonate or use recordings to gain fraudulent access will fail**

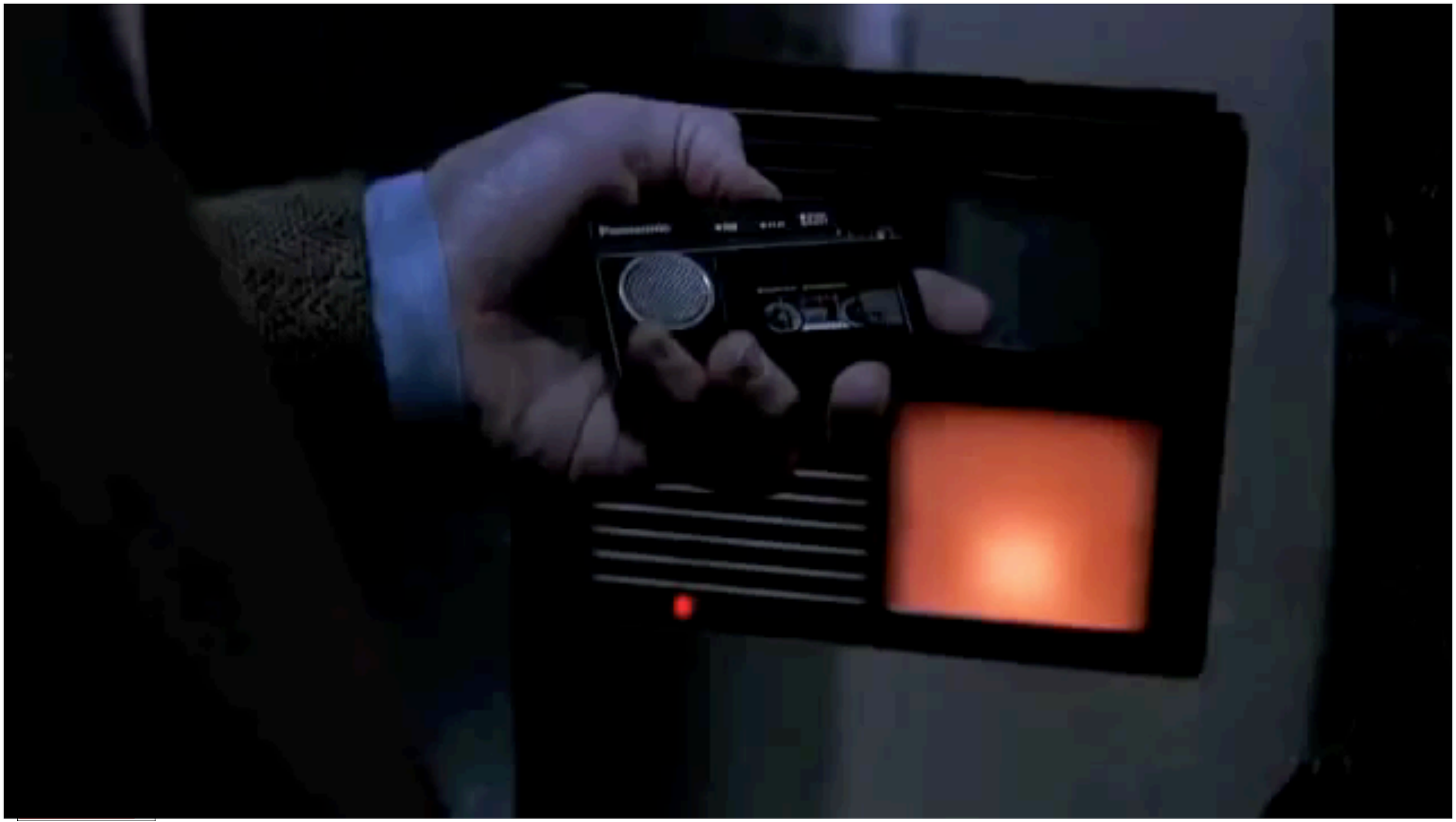
When my voice
is my password...
"I feel secure."

Even under spoofing attacks?

Spoofing attack

Definition: an impostor seeking to deceive the system by impersonating another enrolled user at the microphone in order to manipulate the speaker verification result

An example: **Sneakers** (1992 movie)



About spoofing attack: A recent survey paper tells

Replay spoofing

	No spoofing	Spoofing
Lindberg et al. 1999	1.1 ~ 5.6%	89.5 ~ 100%
Violable et al. 2011	0.71%	68.00%
Wu et al. 2014	2.39 ~ 2.92%	73.14~78.36%

False acceptance rate (FAR)

Speech synthesis spoofing

	No spoofing	Spoofing
Masuko et al. 1999	0.00%	70%
De Leon et al. 2012	0.28%	86%
De Leon et al. 2012	0.00%	81%

Voice conversion spoofing

	No spoofing	Spoofing
Perrot et al. 2005	16%	40%
Kinnunen et al. 2012	3.24%	17.33%
Wu et al. 2012	2.99%	41.25%
Alegre et al. 2013	3.03%	55%

State-of-the-art automatic speaker verification (ASV) systems are vulnerable to spoofing attacks

Zhizheng Wu, Nicholas Evans, Tomi Kinnunen, Junichi Yamagishi, Federico Alegre, Haizhou Li, "Spoofing and countermeasures for speaker verification: a survey", Speech Communication, Volume 66, Pages 130–153, 2015

Will you feel secure?

When my voice is my password…

“I feel secure.”

Spoofing countermeasures (CM): A recent survey paper tells

Replay spoofing

	No spoofing	Spoofing	With CM
Lindberg et al. 1999	1.1 ~ 5.6%	89.5 ~ 100%	n/a
Violable et al. 2011	0.71%	68.00%	0~17%
Wu et al. 2014	2.39 ~ 2.92%	73.14~78.36%	0.00 ~ 0.06%

Speech synthesis spoofing

	No spoofing	Spoofing	With CM
Masuko et al. 1999	0.00%	70%	n/a
De Leon et al. 2012	0.28%	86%	2.5%
De Leon et al. 2012	0.00%	81%	2.5%

Voice conversion spoofing

	No spoofing	Spoofing	With CM
Perrot et al. 2005	16%	40%	n/a
Kinnunen et al. 2012	3.24%	17.33%	n/a
Wu et al. 2012	2.99%	41.25%	1.71%
Alegre et al. 2013	3.03%	55%	4.10

Zhizheng Wu, Nicholas Evans, Tomi Kinnunen, Junichi Yamagishi, Federico Alegre, Haizhou Li, "Spoofing and countermeasures for speaker verification: a survey", Speech Communication, Volume 66, Pages 130–153, 2015

However, lack of standard databases and protocols

- Each study designed its own protocols and database
  - The countermeasure had explicit or implicit prior knowledge of the spoofing attack
  - The generalisation ability of those countermeasures will not be good

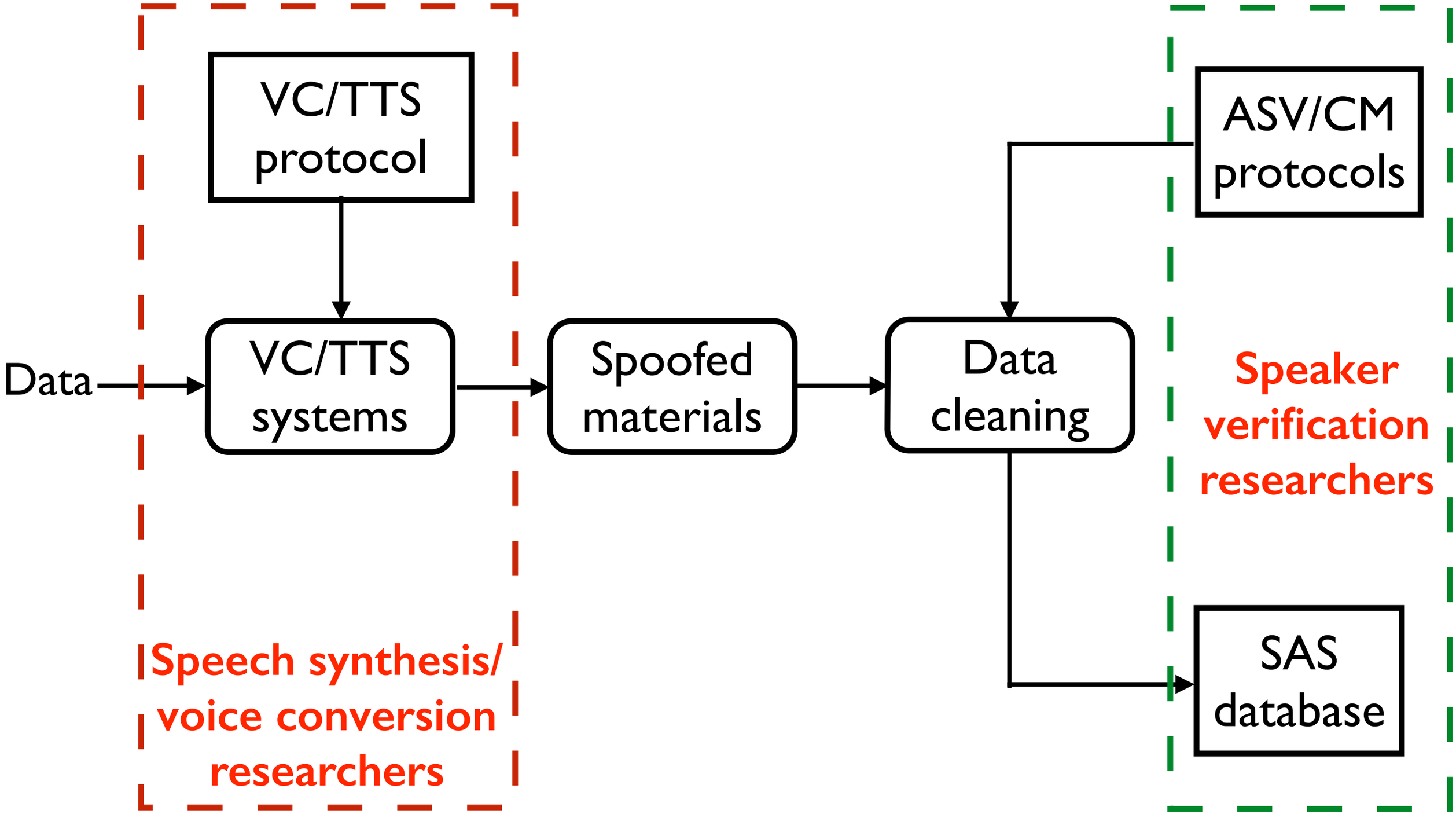
- Review
  - Wall street journal
    - De Leon et al. 2010, De Leon et al. 2012
    - Only speech synthesis
  - NIST Speaker Recognition Evaluation (SRE) 2006
    - Bonastre et al. 2007, Kinnunen et al. 2012, Wu et al. 2012, Alegre et al. 2012, ...
    - Only voice conversion
  - RSR2015
    - Wu et al. 2013, Wu et al. 2014
    - Replay and voice conversion
  - ...

Contributions of this work

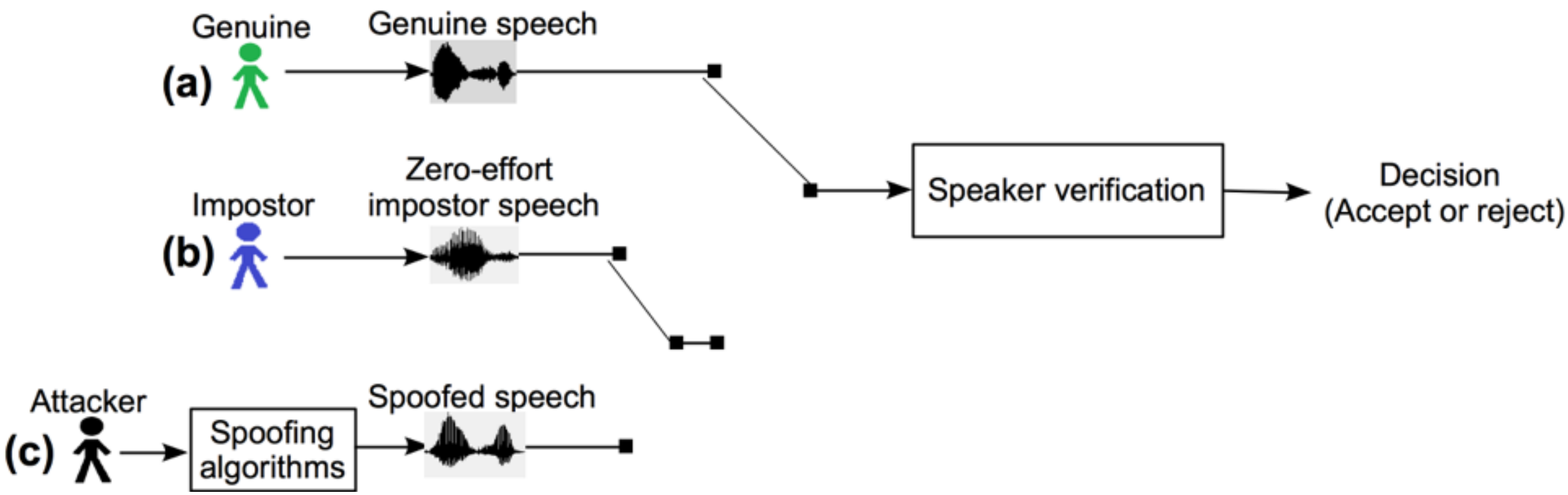
- A spoofing and anti-spoofing (SAS) database with diverse spoofing algorithms
  - 5 speech synthesis algorithms
  - 8 voice conversion algorithms

- Publicly available for free to the community
  - Release under CC-BY license

Flow to generate the SAS database



Framework for generating spoofing materials



(a) + (b): a standard test with genuine and impostor trials

(a) + (c): a spoofing test with genuine and spoofed trials

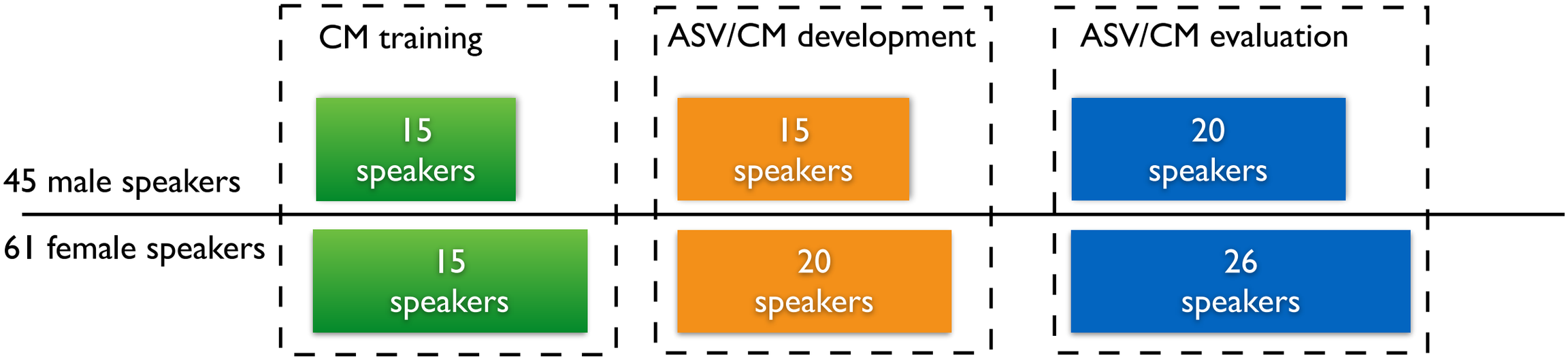
(b) vs (c): same number of trials, same language content

Starting point

- Original corpus: Voice Cloning Toolkit (VCTK) database from the CSTR
  - http://homepages.inf.ed.ac.uk/jyamagis/page3/page58/page58.html
  - 45 male, 61 female
  - Recorded at 96 kHz and down-sampled to 16 kHz in the SAS database
  - No additive channel and noise factors

Speaker and data partition

- **Divide speakers into three subsets**



- **Dividing each speaker's data into 5 parts**
  - **Part A: 24 parallel utterances**
  - **Part B: 20 non-parallel utterances**
  - **Part C: 50 non-parallel utterances**
  - **Part D: 100 non-parallel utterances**
  - **Part E: about 200 non-parallel utterances**

For TTS/VC

For ASV/CM

Spoofing algorithms

- Speech synthesis (Five)
  - SS-LARGE-16
    - HTS, Part A + Part B
  - SS-LARGE-48
    - HTS, Part A + Part B, 48 kHz sampling rate
  - SS-SMALL-16
    - HTS, Part A
  - SS-SMALL-48
    - HTS, Part A, 48 kHz sampling rate
  - SS-MARY
    - MaryTTS system: open-source
    - Waveform concatenation

HTS: http://hts.sp.nitech.ac.jp/
MaryTTS: http://mary.dfki.de/

Spoofing algorithms (cont'd)

- Voice conversion (eight): all use Part A as training data
  - VC-FEST
    - Festvox toolkit: open-source
  - VC-GMM
    - Joint density GMM with dynamic constraints, with some fine-tuning
  - VC-KPLS
    - Kernel partial least square regression
  - VC-EVC
    - Eigenvoice voice conversion
  - VC-TVC
    - Tensor based voice conversion
  - VC-LSP
    - Similar to VC-GMM, but use LSP features
  - VC-FS
    - Frame selection based voice conversion
  - VC-C1
    - Shifting the slope of spectrum

Spoofing algorithms (cont'd)

- **Known attacks** (available for training): 5 algorithms
  - SS-LARGE-16
  - SS-SMALL-16
  - VC-FEST
  - VC-C1
  - VC-FS

- **Unknown attacks** (not available for training): 8 algorithms
  - all the others except known attacks

Protocols

- **ASV protocol**

	Development		Evaluation	
	Male	Female	Male	Female
Target speakers	15	20	20	26
Genuine trials	1498	1999	4053	5351
Impostor trials	12981	17462	32833	46736
Spoofed trials	12981 x 5	17462 x 5	32833 x 13	46736 x 13

- **Countermeasure protocol**

	# Speakers		# Trials	
	Male	Female	Genuine/Human	Spoofed
Training	10	15	3750	12625
Development	15	20	3497	(12981+17462) x 5 = 152215
Evaluation	20	26	9404	(32833 + 46736) x 13 = 1034397

Initial experiments

- Speaker verification systems
  - JFA-5: Joint factor analysis (JFA) with 5-utterance enrolment data
  - JFA-50: JFA with 50-utterance enrolment data
  - PLDA-5: Probabilistic linear discriminative analysis (PLDA) with 5-utterance enrolment data
  - PLDA-50: PLDA with 50-utterance enrolment data

- Background data
  - WSJ0 + WSJ1, WSJ-CAM, Resource Management (RM1)

- Features
  - 19-dimensional MFCC and one energy feature with delta and delta-delta features
  - excluding the static energy feature
  - In total 59 dimension

Spoofing performance: false acceptance rates (FARs)

Spoofing	Male				Female			
	JFA-5	JFA-50	PLDA-5	PLDA-50	JFA-5	JFA-50	PLDA-5	PLDA-50
Baseline	2.76	1.25	1.41	1.16	6.24	2.47	1.52	0.99
SS-LARGE-16	88.62	96.17	93.45	97.76	84.31	84.65	86.04	95.95
SS-LARGE-48	97.62	98.93	99.12	99.09	90.58	94.28	94.80	98.39
SS-MARY	91.09	96.81	96.77	98.74	91.37	95.11	95.28	98.10
SS-SMALL-16	83.64	91.25	89.21	94.87	80.60	77.97	81.60	93.14
SS-SMALL-48	94.97	95.75	97.07	96.63	86.46	90.36	93.02	96.86
VC-C1	2.62	1.46	1.83	1.67	12.71	6.80	3.92	3.56
VC-EVC	43.38	58.52	69.84	79.50	67.82	66.50	72.14	79.10
VC-FEST	91.25	97.71	97.41	99.54	85.77	91.76	86.11	93.53
VC-FS	78.68	91.62	91.05	96.16	71.19	75.32	79.37	90.33
VC-GMM	89.14	96.22	95.10	98.70	85.72	92.93	90.57	97.42
VC-KPLS	61.92	82.90	81.17	89.31	70.99	71.72	80.87	86.32
VC-LSP	51.71	68.37	74.99	89.82	65.30	60.27	70.99	75.14
VC-TVC	63.28	78.75	80.20	87.14	70.87	71.41	74.83	82.20

- The decision threshold is set to the Equal Error Rate (EER) point on the development set without spoofing
- SS-LARGE-48 and SS-MARY result relative much higher FARs than other spoofing algorithms
  - SS-LARGE-48 uses training data of a higher sampling rate (48 kHz)
  - SS-MARY directly concatenates target speech waveform segments
- VC-C1 which only shifts spectral slope also increases the FARs a bit

Take home messages

- A standard spoofing and anti-spoofing (SAS) database contains diverse spoofing algorithms (13 algorithms)

- How to get the database?
  - Send an email request to Zhizheng Wu
  - A downloadable link will be included in this webpage: https://wiki.inf.ed.ac.uk/CSTR/SASCorpus
  - Note: the database is larger than 100 Gb

- If you want to contribute spoofing materials (algorithms), feel free to contact me!

An anti-spoofing challenge has been organised by using a subset of this database
http://www.spoofingchallenge.org/

Email: zhizheng.wu@ed.ac.uk (Zhizheng WU)

Ongoing work

- More spoofing algorithms
  - Consider channel and noise factors
  - More types of waveform concatenation spoofing algorithms
- Phase-preserving vocoders
- Replay attacks
- Text-dependent speaker verification

http://dx.doi.org/10.7488/ds/252

Thank you !