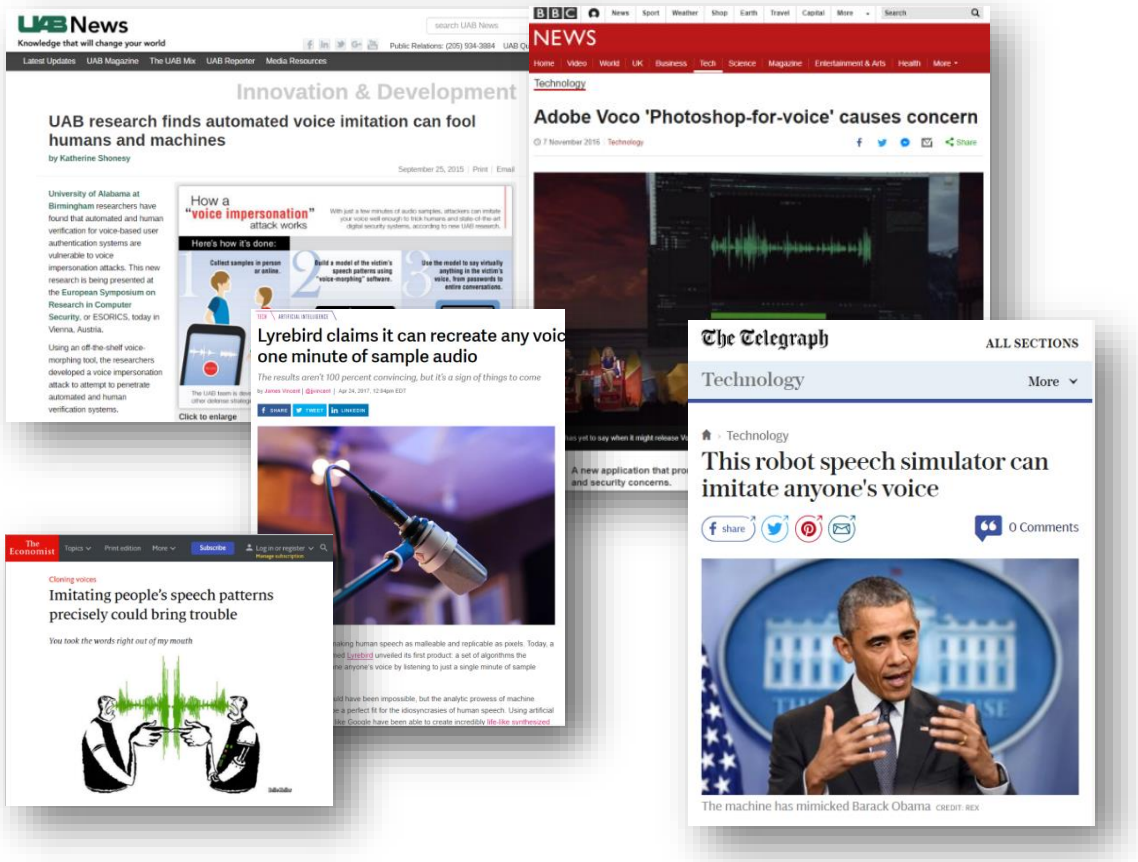
t-DCF:
Detection Cost Function for the  Assessment of Spoofing Countermeasures and Automatic Speaker Verification

Tomi Kinnunen, Kong Aik Lee, Hector Delgado,
Nicholas Evans, Massimiliano Todisco, Md. Sahidullah,
Junichi Yamagishi, Douglas A. Reynolds

Biometric spoofing attacks (presentation attacks)

- ISO standard: ISO/IEC 30107-1:2016
- Ratha, N.K., Connell, J.H., Bolle, R.M., 2001. Enhancing security and privacy in biometrics-based authentication systems. *IBM Systems Journal* 40, 614– 634.
- Z. Wu, N. Evans, T. Kinnunen, J. Yamagishi, F. Alegre, H. Li, "Spoofing and Countermeasures for Speaker Verification: a Survey", *Speech Comm*, 66: 130--153, 2015

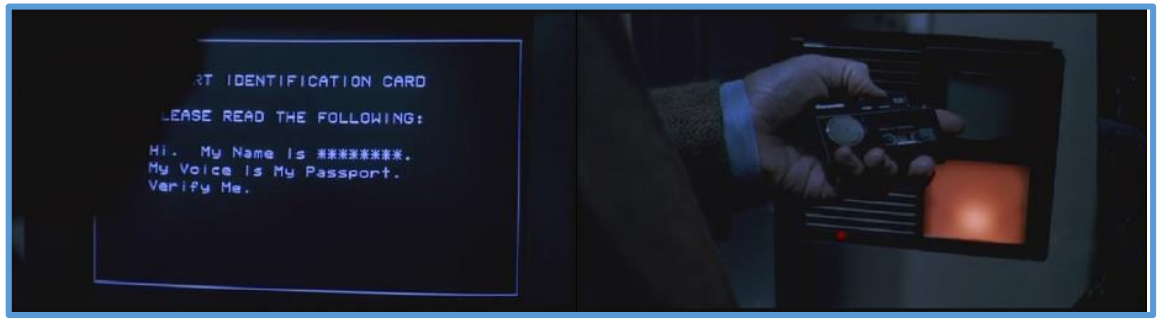
TEXT-TO-SPEECH AND VOICE CONVERSION

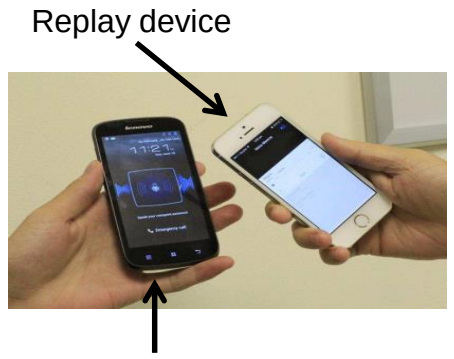


REPLAY

Sneakers (1992)

Universal Pictures





A short history of ASV spoofing research

Special session at Interspeech 2013

OCTAVE project 2015-2017

1999 — 2006 — 2014 — 2016 — 2017 — 2018-2019

SMALL, SELF-COLLECTED DATASETS, CLEAN DATA

STANDARD DATASETS, CLEAN DATA

COMMON DATA, METRICS, PROTOCOLS, CLEAN DATA

OCTAVE
Objective Control of Talker Verification

COMMON DATASETS, REPLAY, GENERALIZATION, CHANNEL AND NOISE VARIATION

ASVspoof 2015

Speech synthesis and voice conversion attacks

ASVspoof 2017

Replay attacks

ASVspoof 2019
(PLANNED)

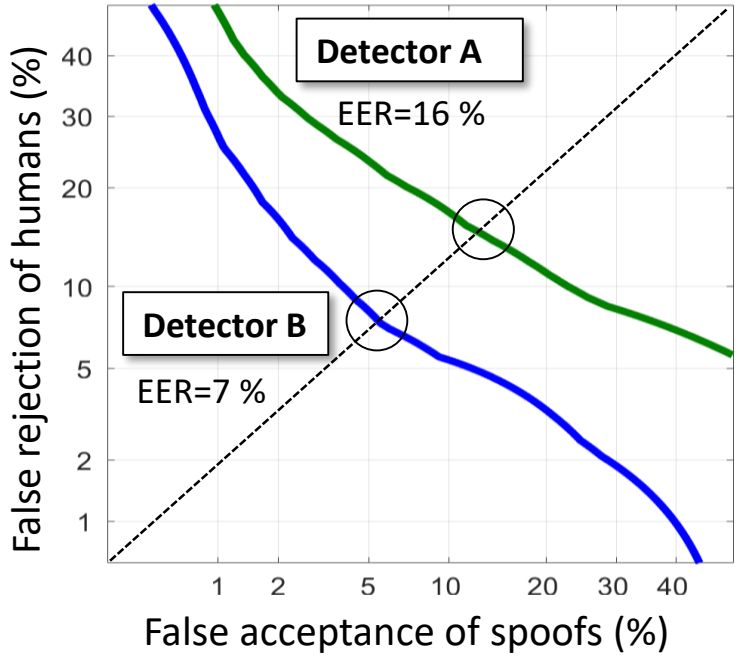
- Synthetic and replay attacks
- ASV-centered evaluation

ASVspoof challenge task: detection of spoofs in isolation from ASV

http://www.asvspoof.org/

ASVspoof 2015

ASVspoof 2017



EVALUATION METRIC IN BOTH CHALLENGES:
EQUAL ERROR RATE (EER) OF THE DETECTOR

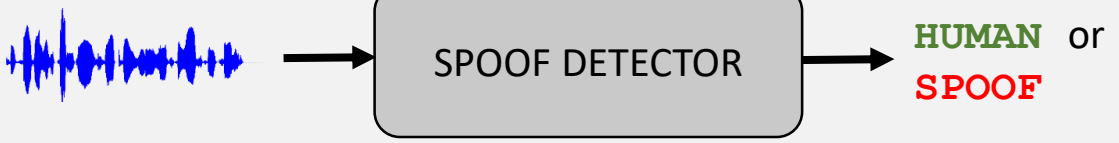
Stockholm, August 2017

ASVspoof 2017 challenge special session
at Interspeech

Hmm, none of this is about ASV ...

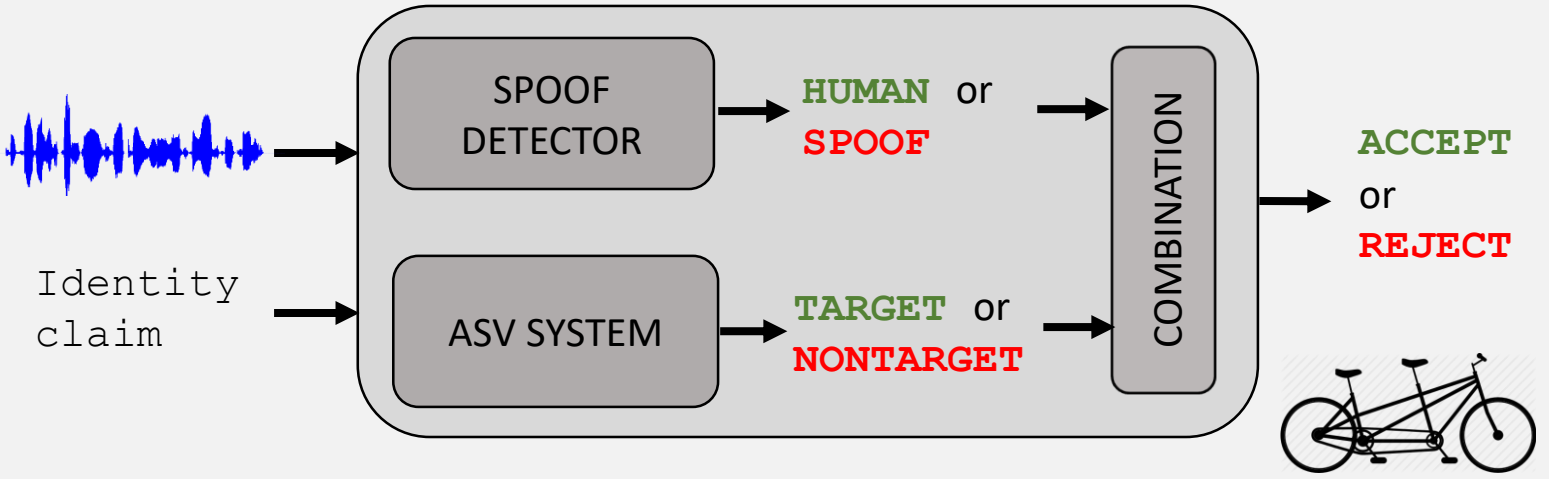
DOUG

Spoof detector in isolation from ASV



SIMPLE - BUT DOES IT IMPROVE ASV ?

ASV system with a spoof detector



A COMPLETE SYSTEM - BUT HOW TO DESIGN AND EVALUATE ?

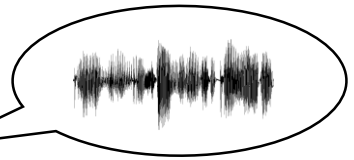
- Combine in series or parallel ?
- The subsystems address different tasks - how do the errors combine ?

Our vision: a metric that

- reflects the performance of the combined (tandem) system
- while remaining backward compatible with standalone assessment
- allows specifying an **application** (costs, user priors)
- facilitates unified comparison of
  - ASV without countermeasure
  - ASV with perfect countermeasure
  - Perfect ASV system with a countermeasure
- is easy to understand and use

The two users in traditional ASV (no spoofing)

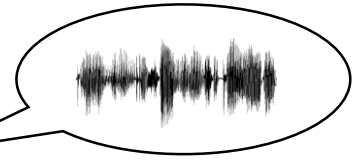
1. TARGET
Should be accepted



Identity claim:
Doug

Detection score

2. NONTARGET
(ZERO-EFFORT IMPOSTOR)
Should be rejected



Identity claim:
Doug

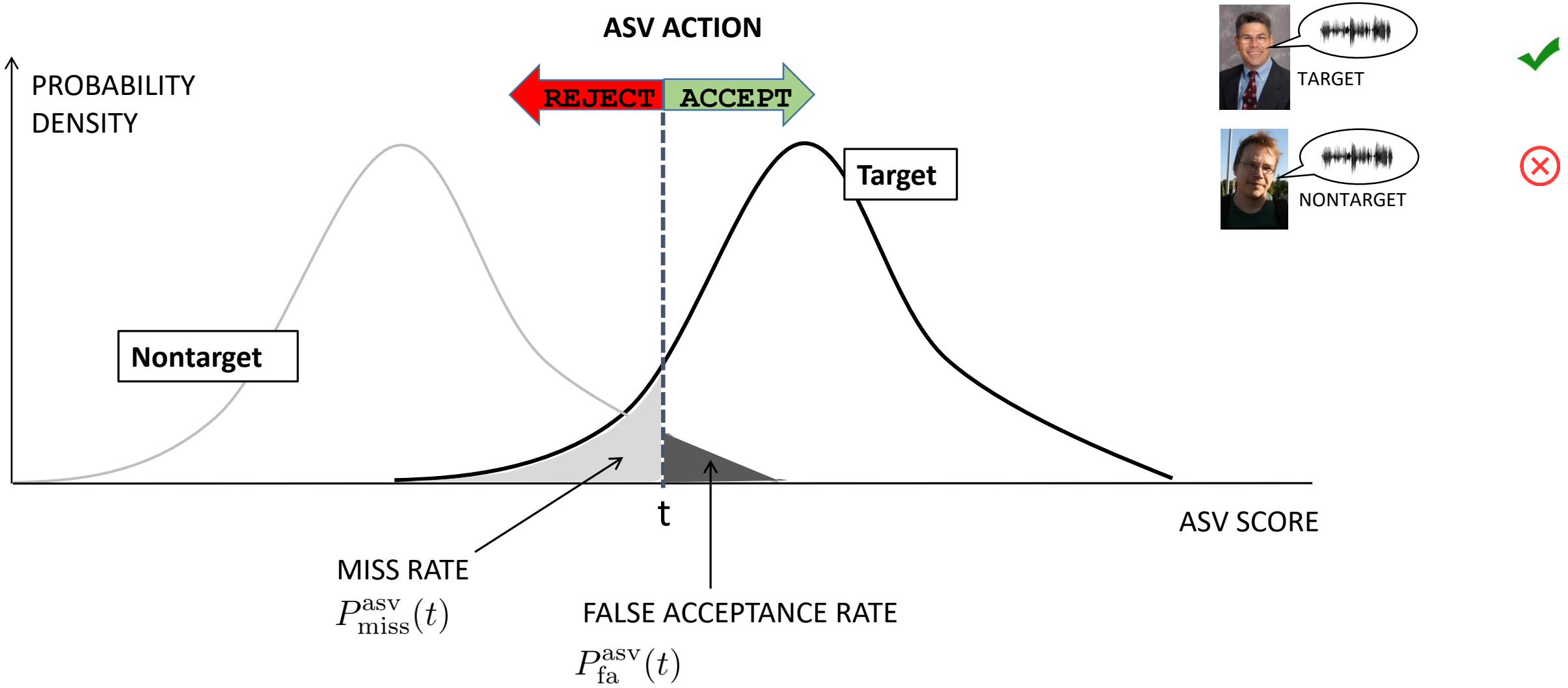
Detection score

Prior probability of each user type

$$\pi_{\text{tar}} = P(\text{target})$$
$$\pi_{\text{non}} = P(\text{nontarget})$$

Note: $\pi_{\text{tar}} + \pi_{\text{non}} = 1$

Error computation in traditional ASV



NIST Detection Cost Function (DCF)

DECIDED IN ADVANCE

COMPUTED FROM EVAL TRIALS

$$\mathrm{DCF}(t) = C_{\mathrm{miss}}^{\mathrm{asv}} \pi_{\mathrm{tar}} P_{\mathrm{miss}}^{\mathrm{asv}}(t) + C_{\mathrm{fa}}^{\mathrm{asv}} (1 - \pi_{\mathrm{tar}}) P_{\mathrm{fa}}^{\mathrm{asv}}(t)$$

COST OF TARGET MISS

TARGET PRIOR

MISS RATE AT THRESHOLD t

COST OF FALSE ACCEPT

NONTARGET PRIOR

FALSE ACCEPTANCE RATE AT THRESHOLD t

TYPICAL NIST EVALUATION

$$C_{\mathrm{miss}}^{\mathrm{asv}} = 1$$
$$C_{\mathrm{fa}}^{\mathrm{asv}} = 1$$
$$\pi_{\mathrm{tar}} = 0.001$$

LOW TARGET USER PRIOR

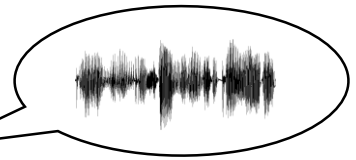
- Surveillance
- Indexing / multimedia search search

The ~~two~~ three types of ASV trial users

Prior probabilities of user types

$$\pi_{\text{tar}} + \pi_{\text{non}} + \pi_{\text{spoof}} = 1$$

1. TARGET
Should be accepted



Identity claim: Doug

ASV SYSTEM

Detection scores

2. NONTARGET (ZERO-EFFORT IMPOSTOR)
Should be rejected

Identity claim: Doug

ASV SYSTEM

Detection scores

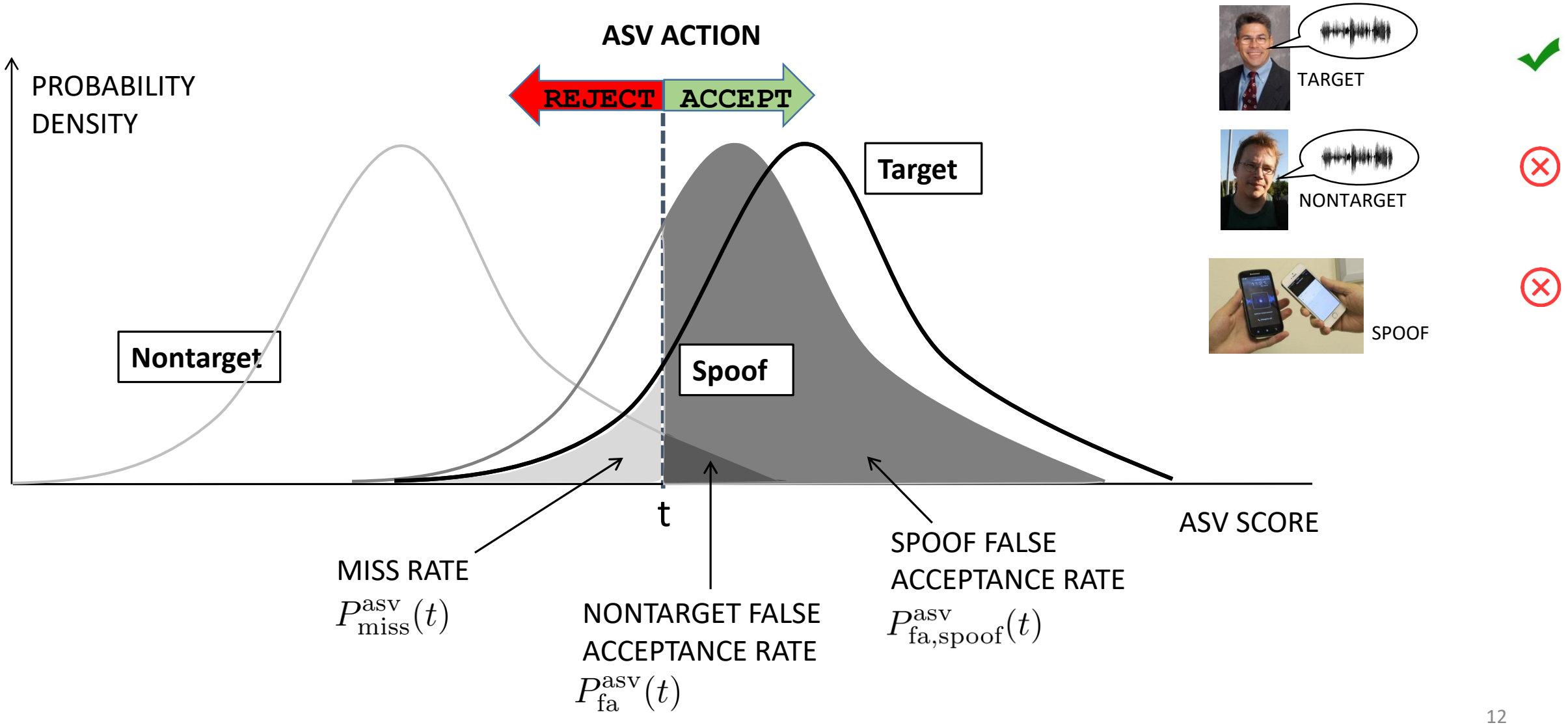
3. SPOOF
Should be rejected

Identity claim: Doug

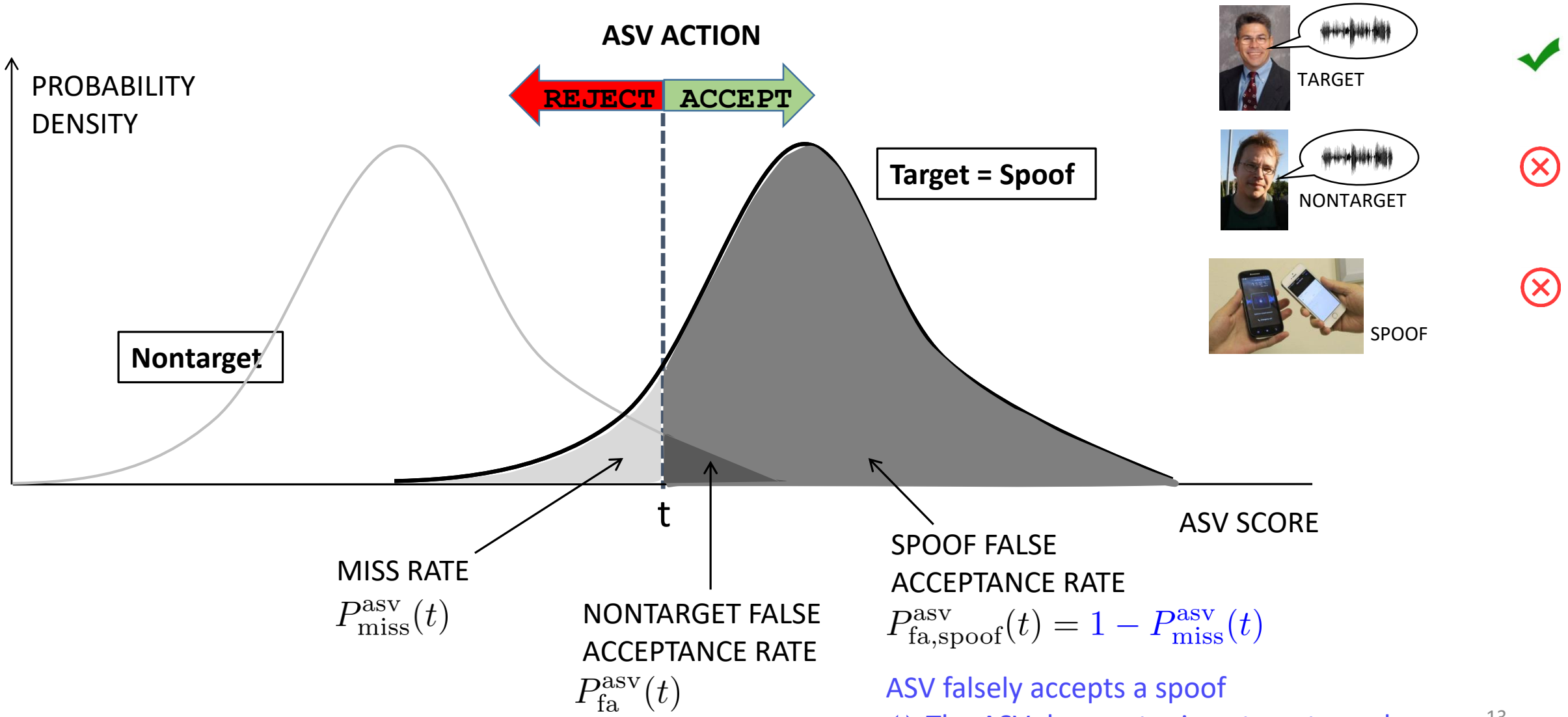
ASV SYSTEM

Detection scores

Automatic speaker verification scores

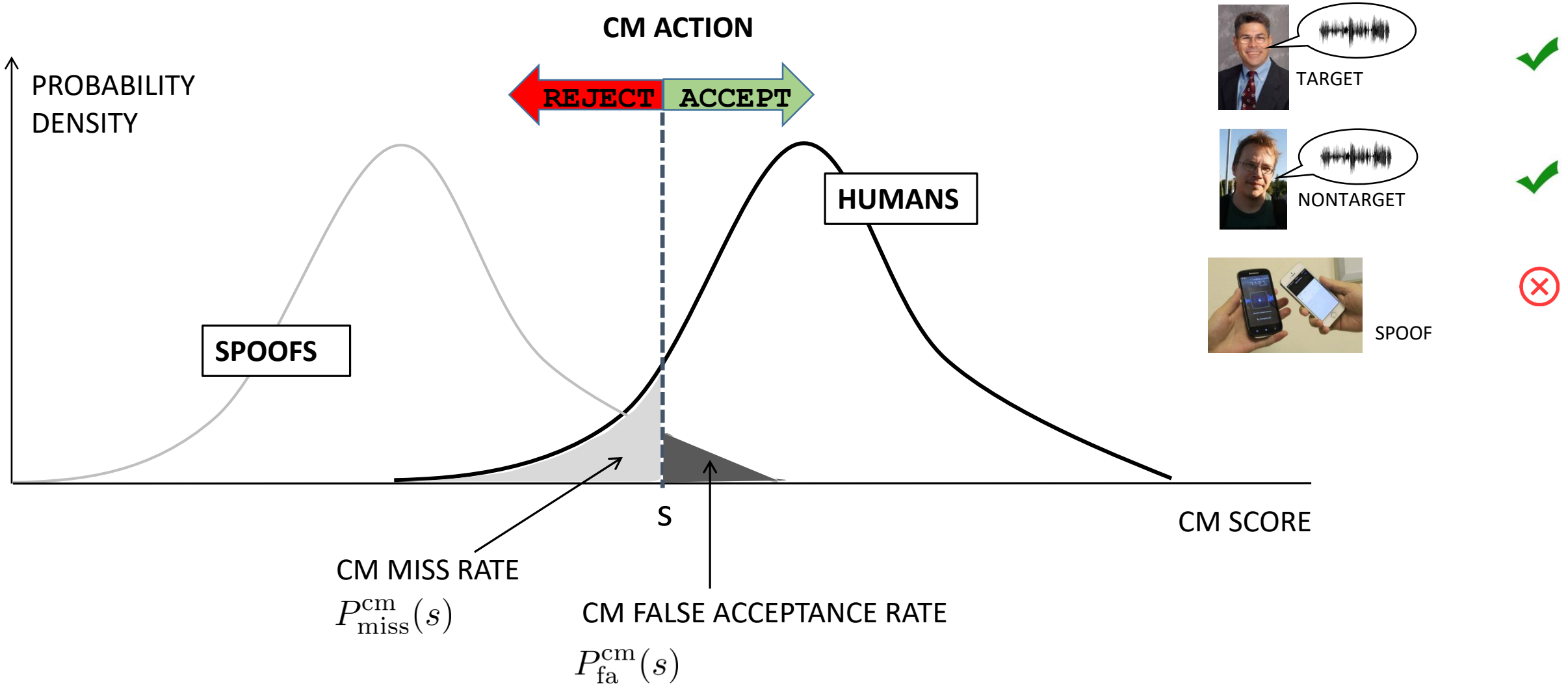


Worst-case scenario: perfect spoofs

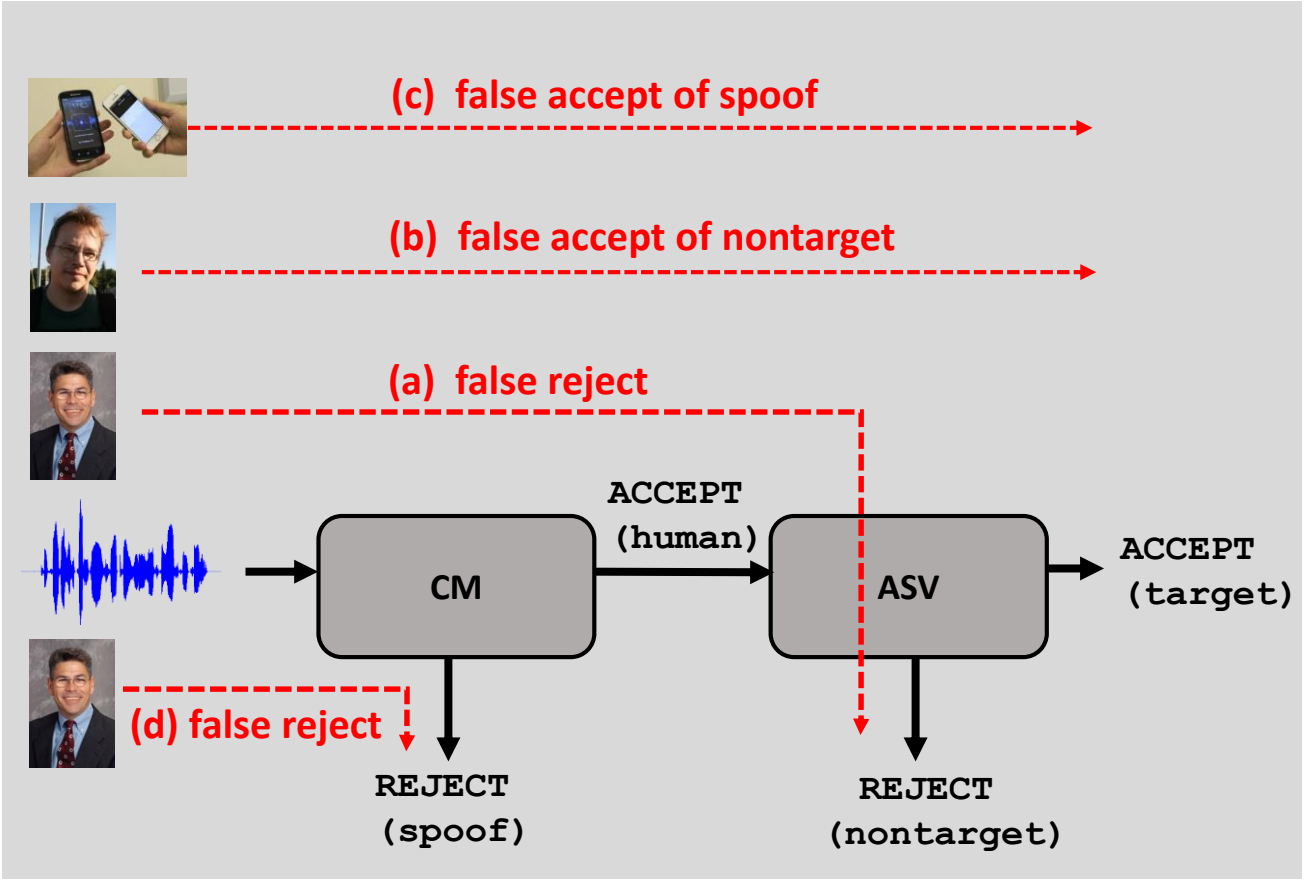


ASV falsely accepts a spoof
⇔ The ASV does not miss a target speaker

Countermeasure scores



The four errors in a tandem system



ERROR PROBABILITIES

(a) *CM does <u>not</u> miss human, and ASV rejects the target.*

$$P_{\mathrm{a}}(s,t) = (1 - P_{\mathrm{miss}}^{\mathrm{cm}}(s)) \times P_{\mathrm{miss}}^{\mathrm{asv}}(t)$$

(b) *CM does <u>not</u> miss human, and ASV falsely accepts nontarget.*

$$P_{\mathrm{b}}(s,t) = (1 - P_{\mathrm{miss}}^{\mathrm{cm}}(s)) \times P_{\mathrm{fa}}^{\mathrm{asv}}(t)$$

(c) *CM falsely accepts spoof, and and ASV does <u>not</u> miss the target*

$$P_{\mathrm{c}}(s,t) = P_{\mathrm{fa}}^{\mathrm{cm}}(s) \times (1 - P_{\mathrm{miss}}^{\mathrm{asv}}(t))$$

(d) *Spoof detector misses human.*

$$P_{\mathrm{d}}(s) = P_{\mathrm{miss}}^{\mathrm{cm}}(s)$$

Tandem Detection Cost Function (t-DCF)

$$\text{t-DCF}(s,t) = C_{\text{miss}}^{\text{asv}} \cdot \pi_{\text{tar}} \cdot P_{\text{a}}(s,t) + C_{\text{fa}}^{\text{asv}} \cdot \pi_{\text{non}} \cdot P_{\text{b}}(s,t) + C_{\text{fa}}^{\text{cm}} \cdot \pi_{\text{spoof}} \cdot P_{\text{c}}(s,t) + C_{\text{miss}}^{\text{cm}} \cdot \pi_{\text{tar}} \cdot P_{\text{d}}(s).$$

DECIDED IN ADVANCE

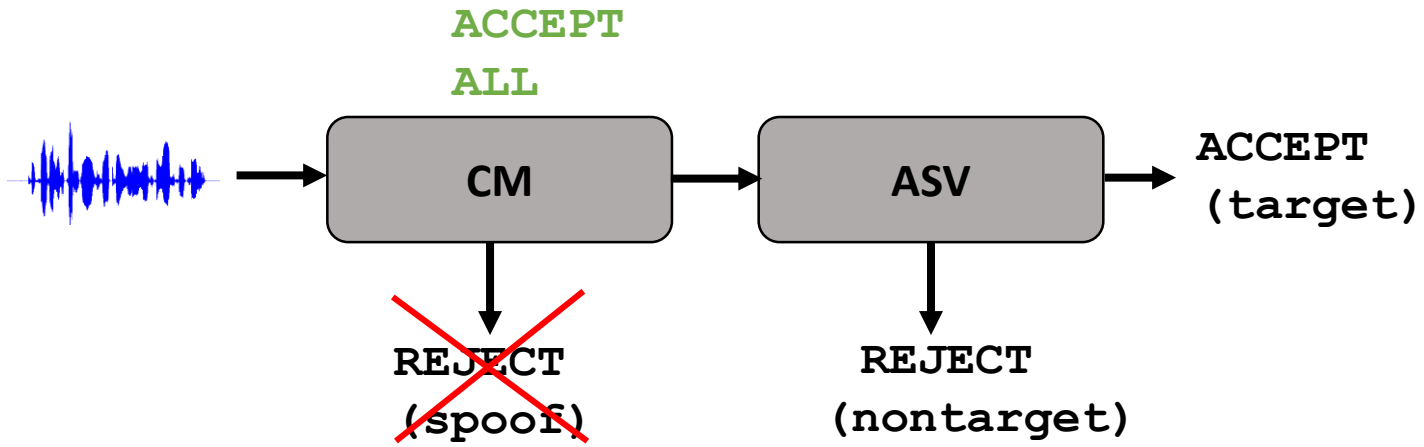
COMPUTED FROM EVAL TRIALS

UNIT COSTS	
$C_{\text{miss}}^{\text{asv}}$	ASV miss (target reject)
$C_{\text{fa}}^{\text{asv}}$	ASV false accept (nontarget accept)
$C_{\text{fa}}^{\text{cm}}$	CM false accept (spoof → human)
$C_{\text{miss}}^{\text{cm}}$	CM miss (human → spoof)

PRIOR OF USER TYPE	
π_{tar}	Target
π_{non}	Nontarget
π_{spoof}	Spoof

$$\pi_{\text{tar}} + \pi_{\text{non}} + \pi_{\text{spoof}} = 1$$

”Accept all” Countermeasure



$$\text{t-DCF}_{\texttt{ACCEPT-ALL}}(t) = C_{\text{miss}}^{\text{asv}} \cdot \pi_{\text{tar}} \cdot P_{\text{miss}}^{\text{asv}}(t) + C_{\text{fa}}^{\text{asv}} \cdot \pi_{\text{non}} \cdot P_{\text{fa}}^{\text{asv}}(t) + C_{\text{fa}}^{\text{cm}} \cdot \pi_{\text{spoof}} \cdot (1 - P_{\text{miss}}^{\text{asv}}(t))$$

$$\text{DCF}(t) = C_{\text{miss}}^{\text{asv}} \pi_{\text{tar}} P_{\text{miss}}^{\text{asv}}(t) + C_{\text{fa}}^{\text{asv}} (1 - \pi_{\text{tar}}) P_{\text{fa}}^{\text{asv}}(t)$$

FURTHER ASSUMPTION: NO SPOOFS

$$\pi_{\text{spoof}} = 0$$

NIST DCF

Other special cases

COUNTERMEASURE THAT REJECTS EVERY INPUT

$$\text{t-DCF}_{\texttt{REJECT-ALL}} = C_{\text{miss}}^{\text{cm}} \pi_{\text{tar}}$$

NO COSTS FROM NONTARGETS OR SPOOFS – THEY ARE ANYWAY REJECTED

ORACLE (UPPER BOUND) FOR A FIXED ASV SYSTEM

$$\begin{aligned}\text{t-DCF}_{\texttt{IDEAL-CM}}(t) = {} & C_{\text{miss}}^{\text{asv}} \cdot \pi_{\text{tar}} \cdot P_{\text{miss}}^{\text{asv}}(t) \\ & + C_{\text{fa}}^{\text{asv}} \cdot \pi_{\text{non}} \cdot P_{\text{fa}}^{\text{asv}}(t)\end{aligned} \qquad \pi_{\text{tar}} + \pi_{\text{non}} \neq 1$$

t-DCF for a ”bank” application

PRIORS:

- Fix π_{spoof} to a low number, e.g. $\pi_{\text{spoof}} = 0.001$
- Set $\pi_{\text{tar}} = (1 - \pi_{\text{spoof}}) \times 0.99$
- Set $\pi_{\text{non}} = (1 - \pi_{\text{spoof}}) \times 0.01$

LOW SPOOF PRIOR

HIGH TARGET PRIOR

LOW NONTARGET PRIOR

COSTS:

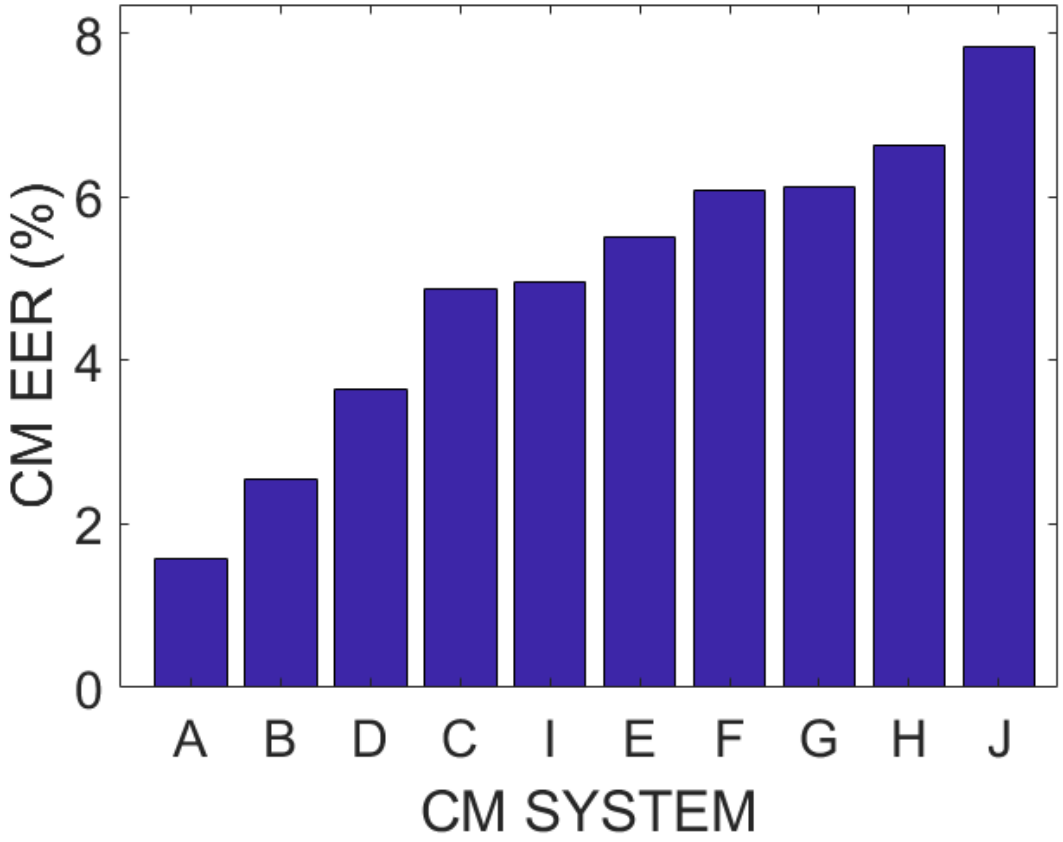
- Set $C_{\text{fa}}^{\text{asv}} = C_{\text{fa}}^{\text{cm}} = 10$
- Set $C_{\text{miss}}^{\text{asv}} = C_{\text{miss}}^{\text{cm}} = 1$

FALSE ACCEPTANCE IN
EITHER SYSTEM MORE COSTLY

Top-10 countermeasures from ASVspoof

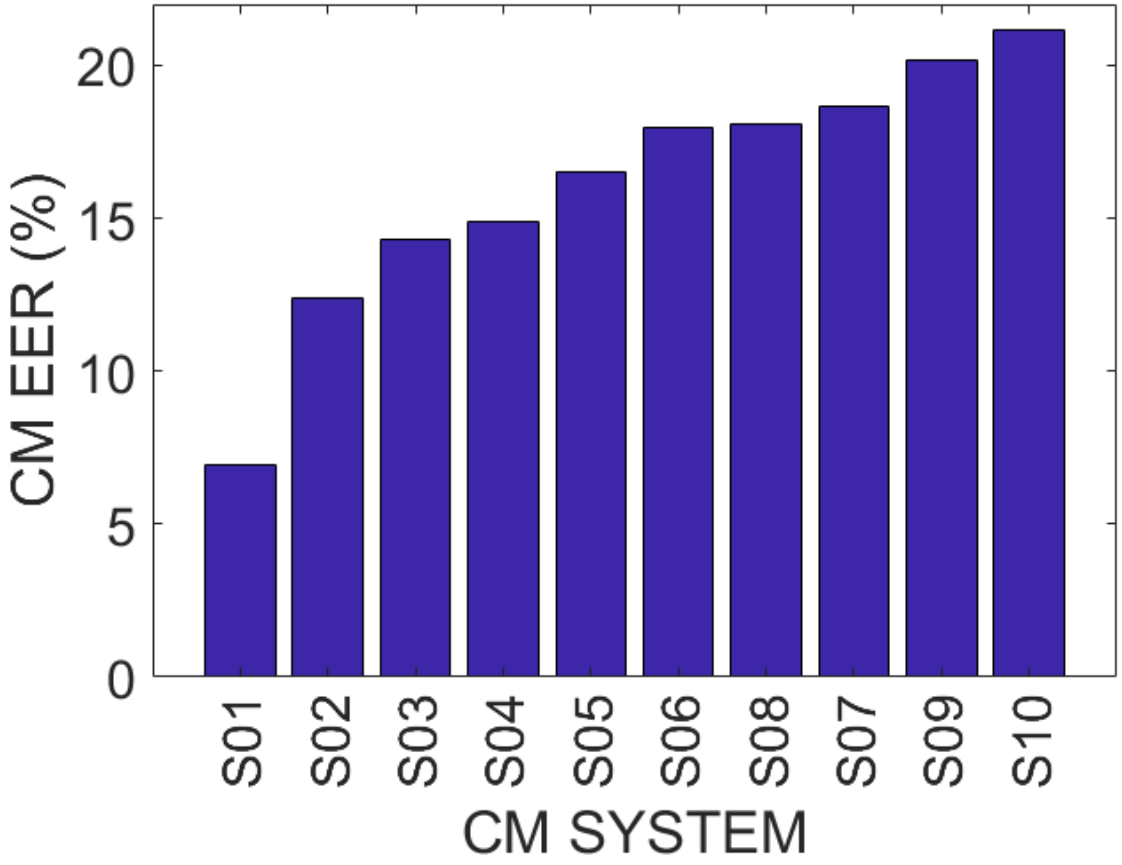
ASVspoof 2015

TTS & VC attacks, high quality audio

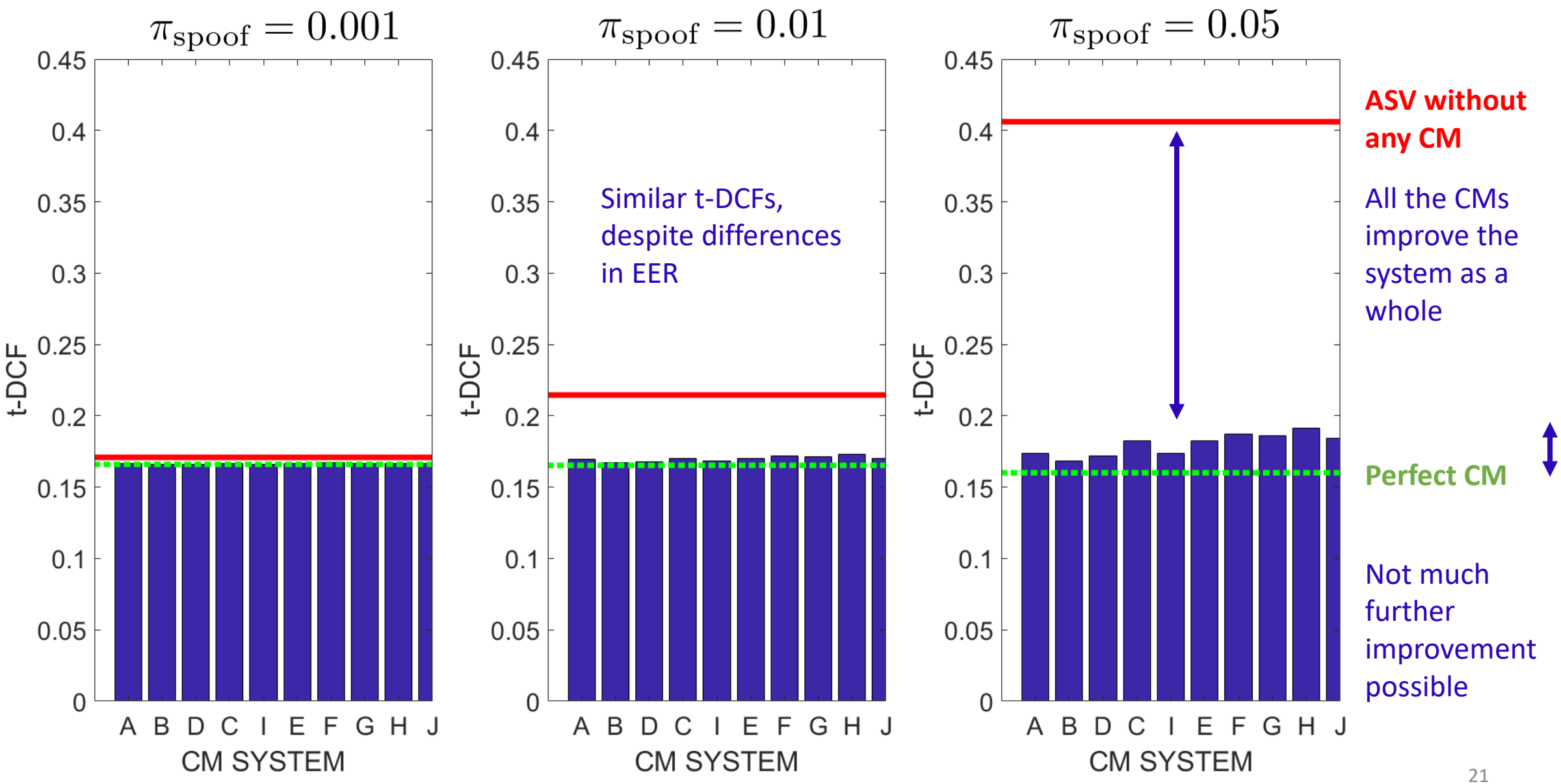


ASVspoof 2017

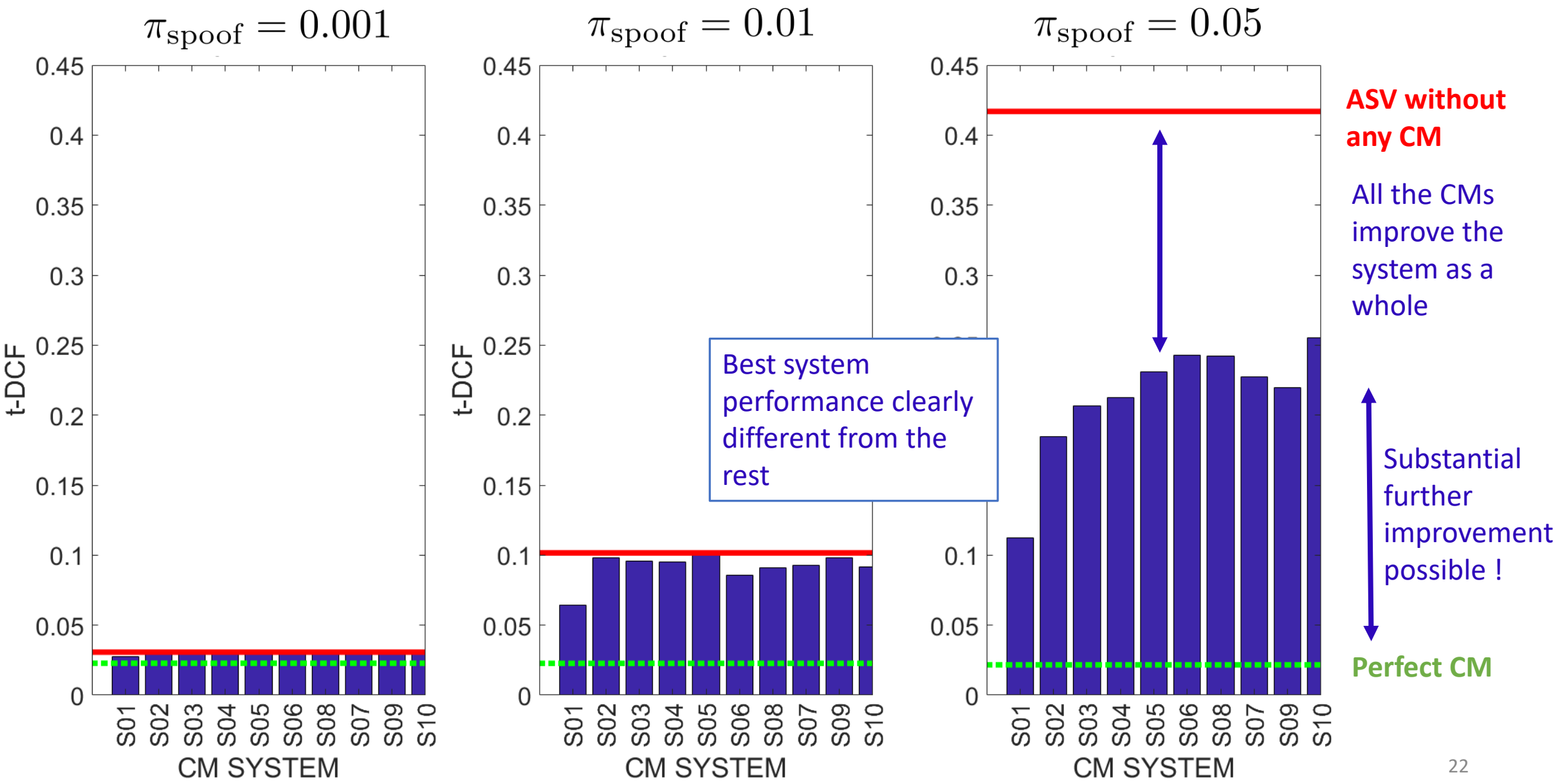
Replay attacks, noisy/varied audio



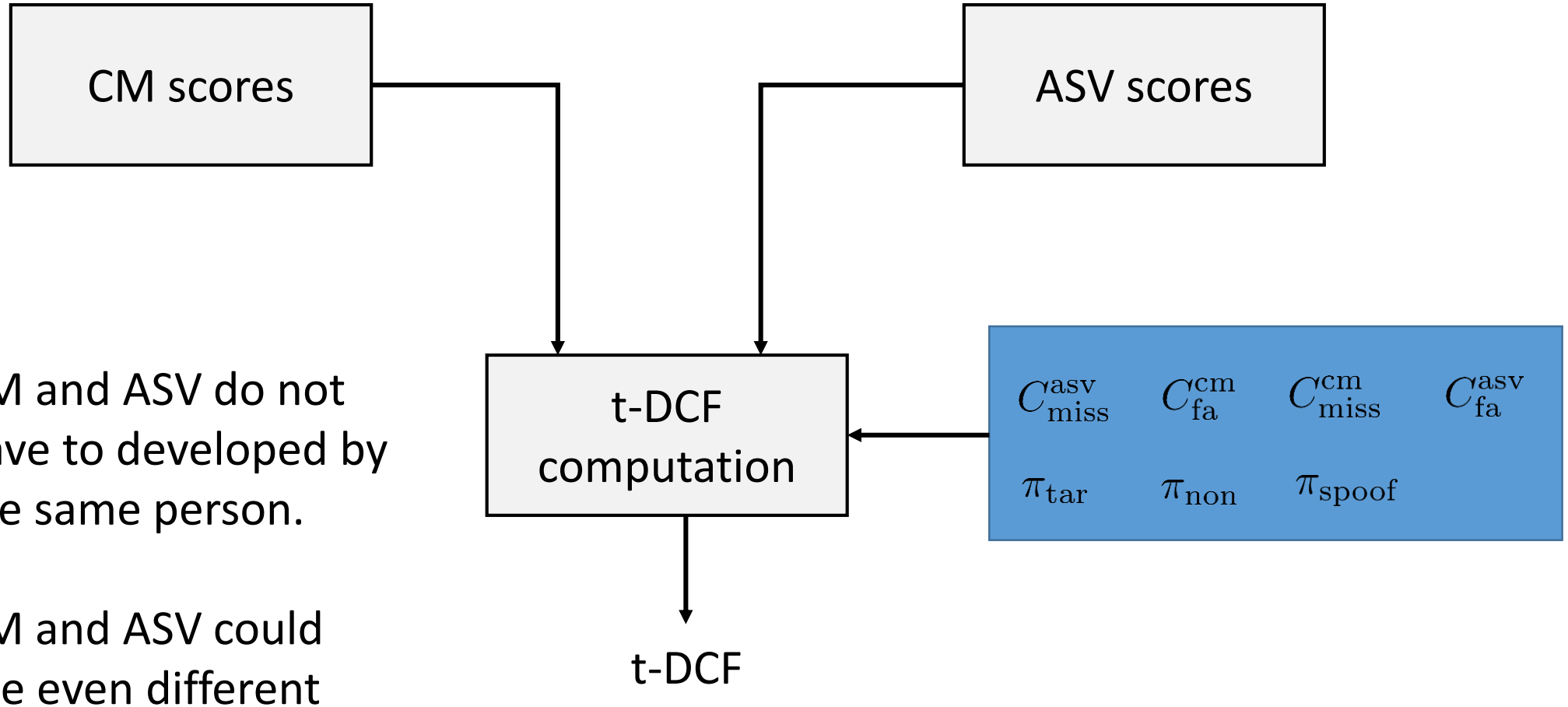
t-DCF, ASVspoof 2015



t-DCF, ASVspoof 2017



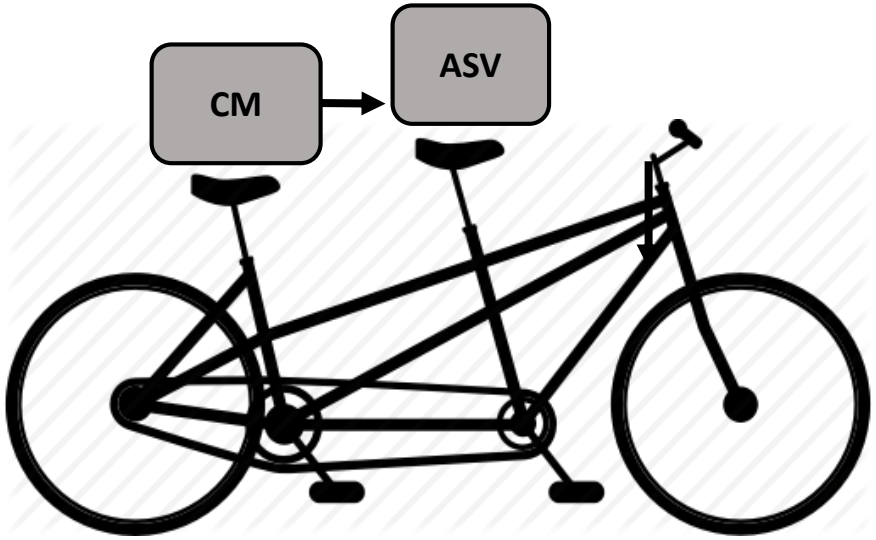
Simplicity of evaluation is retained



- CM and ASV do not have to developed by the same person.
- CM and ASV could use even different evaluation corpora.

Conclusion

- Explicit cost function model for tandem ASV and CM evaluation
- NIST DCF is a special case of t-DCF
- Planned as the primary metric of ASVspoof 2019



http://www.asvspoof.org/data2017/tDCF_v0.1.zip

ASV system

- **ASV system:** GMM-UBM, UBM from TIMIT
- **Features:** 19 static MFCC + Δ + $\Delta^{2,}$ RASTA, energy SAD, CMVN

Table 2: Number of trials in the ASVspoof 2015 and ASVspoof 2017 evaluation protocols for ASV experiments.

Trial Type	ASVspoof 2015	ASVspoof 2017
Target	4053	1106
Nontarget	77007	18624
Spoof	80000	10878