Speech Technology Center

PhoneSpoof: A new dataset for spoofing attack detection in telephone channel

ITMO UNIVERSITY

Galina Lavrentyeva[1,2] , Sergey Novoselov[1,2] , Marina Volkova[2] , Yuri Matveev[1,2] , Maria De Marsico[3]

[1]ITMO University, St. Petersburg, Russia; [2]STC-innovations Ltd., St. Petersubrg, Russia; [3]Sapienza University of Rome, Rome, Italy

{lavrentyeva, novoselov, volkova, matveey}@speechpro.com, demarsico@di.unroma1.it

1 Introduction

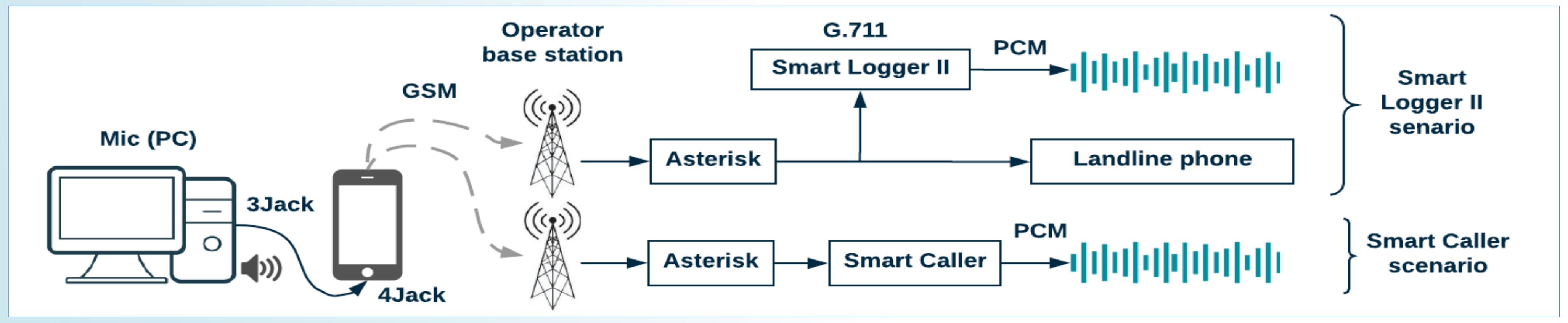
- Automatic Speaker Verification (ASV) remains vulnerable to **spoofing attacks**
- **ASVspoof initiative** [1,2] has significantly pushed forward the development of spoofing detection methods for ASV systems in microphone channel
- **Telephone channel** presents much more challenging conditions for spoofing detection, due to limited bandwidth, various coding standards and channel effects
- Research on the topic has thus far only made use of program codecs and other telephone channel emulations [3]
- In order to asses spoofing detection methods in real scenario we present the **PHONESPOOF dataset** - spoofing data collected through **realistic telephone channels**

3 Emulated telephone conditions

Channel simulation strategies

- **Emulation condition E1:** Original sample rate down sampling to 8kHz and software codec G.6.10 implementation with 13kbit/s bitrate to emulate lossy speech compression in cellular telephony without package loss
- **Emulation condition E2**: STC-H219 [4] sound device for analogue signals recording sent through 2 meters telephone cable

4 PHONESPOOF: real telephone spoofing data



The real spoofing database

- was collected by recording over the telephone channel items from the two datasets of ASVspoof challenges
- includes the subsets of Text-to-Speech (TTS) samples created by cloud services and available libraries: Google , Yandex , IBM , Lyrebird, Zamzar , Ispeech and STC
- collected using ”Smart Logger II” (Recording R1) sound system for telephone calls and speech messages registration
- collected using ”Smart Caller” (Recording R2) system of voice notification via telephone lines

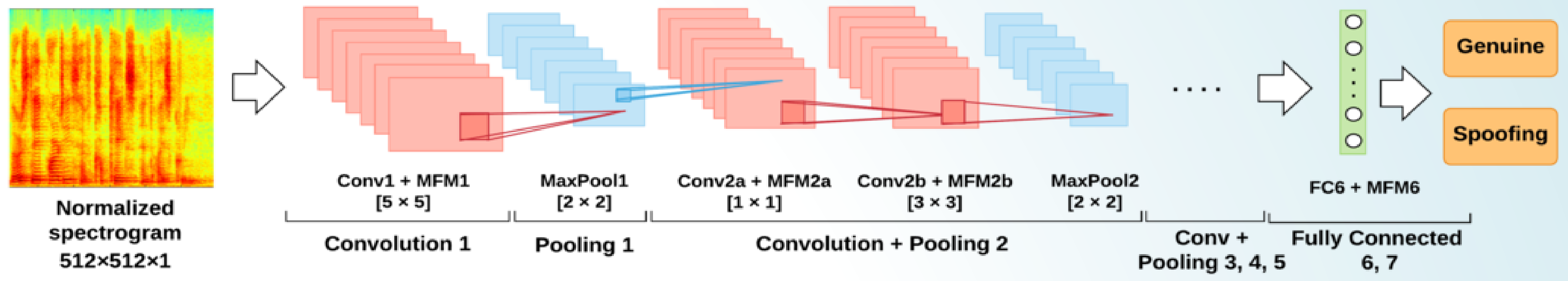
Table 1. Total duration of collected spoofing trials in microphone and telephone channels in hours

Dataset	Language	Microphone channel	Recorded R1	Recorded R2
ASVspoof2015$_{sp}$	eng	177.1	361.6	856.6
ASVspoof2015$_{g}$	eng	160.2	51.2	-
iSpeech	eng	5.47	-	-
IBM	eng	203.4	-	19.9
	rus	217.0	-	15.1
Zamzar	eng	210.8	-	-
STC	rus	523.1	-	48.4
Yandex	eng	236.3	6.2	53.4
	rus	201.7	3.3	56.8
Google	eng	314.8	6.6	46.0
	rus	240.6	3.3	48.7
Lyrebird	eng	95.9	-	1.64
RSR_phrases	eng	150.7	-	29.9
RedDots2015	eng	142.8	-	30

5 Anti-Spoofing system

Anti-spoofing systems under consideration:

- Different CQCC-GMM based anti-spoofing systems for logical (Voice Conversion, TTS) and physical (Replay) spoofing detection
- A unified Light CNN-based approach for both logical and physical spoofing attack detection



6 Experimental results

Table 2 . Experiment results for CQCC-GMM based anti-spoofing system , EER(%)

Emulation type	original	8kHz	6.10 codec
ASVspoof2015	2.24	45.46	46.35

Table 3 . Experiment results for different languages, EER(%)

Training set	Evaluation set	EER (%)
English$_{train}$	Russian$_{eval}$	5.52
	English$_{eval}$	0.03
English$_{train}$ + Russian$_{train}$	Russian$_{eval}$	0.51
	English$_{eval}$	0.14

English$_{train/eval}$ - geniuine : {NIST + ASVspoof$_{g}$};
spoof: ASVspoof$_{sp}$+ Google(eng) + Yandex(eng)
Russian$_{train/eval}$ - geniuine : {RusTelecom};
spoof: ASVspoof$_{sp}$+ Google(rus) + Yandex(rus)

Table 4 . Experiment results for different spoofing types for LCNN, EER(%)

ASVspoof2015 R1				ASVspoof2015 R2		
TTS		VC		TTS		VC
2.74		3.00		0.97		1.27
Google R2		Yandex R2		IBM R2		Replay R2
Eng	Rus	Eng	Rus	Eng	Rus	
1.88	0.86	0.20	1.49	2.45	3.16	1.77

Table 3 . Experiment results for emulated and recorded conditions, EER(%)

Training set	Evaluation set	EER (%)
E1-Emulated$_{train}$	E2-Emulated$_{eval}$	10.98
E2-Emulated$_{train}$		7.79
	R1-Recorded$_{eval}$	26.85
R1-Recorded$_{train}$	E2-Emulated$_{eval}$	49.90

E1-Emulated$_{train}$ - geniuine : {NIST + ASVspoof$_{g}$+ RusTel.} E1; spoof: {ASVspoof$_{sp}$+ Google(eng+rus) + Yandex (eng + rus)} E1
E2-Emulated$_{train}$ - geniuine : {ASVspoof$_{g}$} E2; spoof: {ASVspoof$_{sp}$}E2
E2-Emulated$_{eval}$ - geniuine : {ASVspoof$_{g}$} E2; spoof: {ASVspoof$_{sp}$}E2
R1-Recorded$_{eval}$ - geniuine : {NIST} ; spoof: {ASVspoof$_{sp}$}R1

7 Conclusions

- PHONESPOOF data collection consists of audio spoofing attacks collected through real telephone channels
- Regular telephone channel emulation does not quite match the realistic telephone spoofing attacks scenario which is highly important for the developing of anti-spoofing systems suitable for real applications
- Adding target language to the training set enhance spoofing detection performance on this language
- Efficiency of deep learning frameworks for solving the considered task is confirmed

8 References

1. Z. Wu, T. Kinnunen, N. W. D. Evans, J. Yamagishi, C. Hanili, M. Sahidullah, and A. Sizov, “Asvspoof 2015: the first automatic speaker verification spoofing and countermeasures challenge,” in INTERSPEECH , 2015.
2. T. Kinnunen, M. Sahidullah, H. Delgado, M. Todisco, N. W. D. Evans, J. Yamagishi, and K.-A. Lee, “The asvspoof 2017 challenge: Assessing the limits of replay spoofing attack detection,” in INTERSPEECH , 2017
3. H. Delgado, M. Todisco, N. Evans, M. Sahidullah, W. M. Liu, F. Alegre, T. Kinnunen, and B. Fauve, “Impact of bandwidth and channel variation on presentation attack detection for speaker verification,” BIOSIG 2017
4. “STC H219 overview.” [Online]. Available: http://speechpro.com/product/voice-recording/smartlogger2#tab4

This work was partially financially supported by the Government of the Russian Federation (Grant 08-08).

www.speechpro.com