Post-evaluation analysis for the VoicePrivacy 2020 Challenge: Using anonymized speech data to train attack models and ASR

Natalia Tomashenko[1], Brij Mohan Lal Srivastava[2], Xin Wang[3], Emmanuel Vincent[4], Andreas Nautsch[5], Junichi Yamagishi[3,6], Nicholas Evans[5], Jose Patino[5], Jean-François Bonastre[1], Paul-Gauthier Noé[1], and Massimiliano Todisco[5]

[1]Laboratoire Informatique d'Avignon (LIA), Avignon Université, France
[2]Inria, France
[3]National Institute of Informatics, Tokyo, Japan
[4]Université de Lorraine, CNRS, Inria, LORIA, France
[5]Audio Security and Privacy Group, EURECOM, France
[6]University of Edinburgh, UK

https://voiceprivacychallenge.org

Abstract

This document is related to the post-evaluation analysis for The VoicePrivacy 2020 Challenge [1]. As a part of the post-evaluation analysis, we investigate the anonymized data collection for two different problems: (1) training more advanced attack models and (2) using it in downstream tasks, in particular, for ASR training. We describe the data required from the volunteer participants to perform this analysis and present the results for two baseline anonymization systems.

1 Post-evaluation analysis

Following Deadline-2 of the VoicePrivacy Challenge [2], the organizers will run additional evaluation experiments in order to further characterize the performance of the submitted systems and pave the way for the next Voice Privacy Challenge. To do so, we will ask volunteer participants to share with us the anonymized speech data obtained when running their anonymization system on the ASR/ASV training dataset, and we will compute additional evaluation metrics using these data.

Roughly speaking, this involves:

1. retraining the evaluation systems (ASR_{eval} and ASV_{eval}) on anonymized training data in order to evaluate the suitability of the proposed anonymization technique for ASR and ASV training;

2. computing subjective evaluation metrics on selected subsets of evaluation data;

3. computing additional metrics to assess, e.g., the fulfillment of the goal that pseudo-speakers are sufficiently different from each other and the speaker verifiability performance in the presence of an attacker with additional knowledge.

2 Data submission

The submission of additional data for post-evaluation analysis should include anonymized speech data (wav files, 16kHz, with the same names as in the original corpus) generated from the ***LibriSpeech-train-clean-360*** dataset: http://www.openslr.org/resources/12/train-clean-360.tar.gz.

For experiments, wav files will be converted to 16-bit Signed Integer PCM format, and this format is recommended for submission. All the anonymized speech data should be submitted in the form of a single compressed archive and uploaded to the sftp challenge server voiceprivacychallenge.univ-avignon.fr with a personal login and password, into the subdirectory *postevaluation*. The name of the archive file should correspond to the name of the team used in registration.

Anonymization should be performed in the same way as for the evaluation data, in a speaker-to-speaker manner, using the same anonymization algorithm as for the system submitted as ***primary***.

3 Retraining of ASR_{eval} and ASV_{eval} on anonymized data

The users' downstream goals and the attack models listed in the evaluation plan [2] are not exhaustive. For instance, beyond ASR decoding, anonymization is extremely useful in the context of anonymized data collection for ASR training. It is also known that the EER becomes lower when the attackers generate anonymized training data and retrains ASV_{eval} on this data.

In this section, for the VoicePrivacy anonymization baseline systems ([2], Section 6), we present the objective evaluation results obtained using the ASR_{eval} and ASV_{eval} models trained on the anonymized data. The models were trained as described in [2], but instead of the original speech data we used anonymized data. The anonymization of the training data was performed in the same way as for the development and evaluation data in the speaker-to-speaker manner.

For Baseline-1, Tables 1 and 2 demonstrate ASV and ASR objective evaluation results for the evaluation models trained on the original and anonymized data.

#	ASR_{eval}	**Dev set**	**WER, %**		**Data**	**Test set**	**WER, %**	
			LM$_s$	**LM**$_l$			**LM**$_s$	**LM**$_l$
1	o	libri_dev	5.25	3.83	o	libri_test	5.55	4.15
2	o	libri_dev	8.76	6.39	a	libri_test	9.15	6.73
	a	libri_dev	6.13	4.40	a	libri_test	6.37	4.77
3	o	vctk_dev	14.00	10.79	o	vctk_test	16.39	12.82
4	o	vctk_dev	18.92	15.38	a	vctk_test	18.88	15.23
	a	vctk_dev	14.80	11.67	a	vctk_test	15.00	11.68

Table 2: ***Baseline-1***: ASR decoding error achieved by ASR_{eval} trained on the original data and anonymized data. The results highlighted in blue are the new results obtained using the ASV_{eval} model trained on anonymized speech data, and the other results in the table correspond to the results reported in [2], Table 9.

For Baseline-2, Tables 3 and 4 demonstrate ASV and ASR objective evaluation results for the evaluation models trained on the original and anonymized data.

#	ASV_{eval}	Dev. set	EER,%	$C_{\text{llr}}^{\text{min}}$	C_{llr}	Enroll	Trial	Gender	Test set	EER,%	$C_{\text{llr}}^{\text{min}}$	C_{llr}
1	o	libri_dev	8.665	0.304	42.857	o	o	f	libri_test	7.664	0.183	26.793
2	o	libri_dev	50.140	0.996	144.112	o	a	f	libri_test	47.260	0.995	151.822
3	o	libri_dev	36.790	0.894	16.345	a	a	f	libri_test	32.120	0.839	16.270
	a	libri_dev	18.890	0.563	6.946	a	a	f	libri_test	12.230	0.384	3.004
4	o	libri_dev	1.242	0.034	14.250	o	o	m	libri_test	1.114	0.041	15.303
5	o	libri_dev	57.760	0.999	168.988	o	a	m	libri_test	52.120	0.999	166.658
6	o	libri_dev	34.160	0.867	24.715	a	a	m	libri_test	36.750	0.903	33.928
	a	libri_dev	7.453	0.241	3.585	a	a	m	libri_test	10.690	0.329	5.082
7	o	vctk_dev_com	2.616	0.088	0.868	o	o	f	vctk_test_com	2.890	0.091	0.866
8	o	vctk_dev_com	49.710	0.995	172.049	o	a	f	vctk_test_com	48.270	0.994	162.531
9	o	vctk_dev_com	27.910	0.741	7.205	a	a	f	vctk_test_com	31.210	0.830	9.015
	a	vctk_dev_com	14.530	0.473	1.575	a	a	f	vctk_test_com	18.790	0.552	1.978
10	o	vctk_dev_com	1.425	0.050	1.559	o	o	m	vctk_test_com	1.130	0.036	1.041
11	o	vctk_dev_com	54.990	0.999	192.924	o	a	m	vctk_test_com	53.390	1.000	190.136
12	o	vctk_dev_com	33.330	0.840	23.891	a	a	m	vctk_test_com	31.070	0.835	21.680
	a	vctk_dev_com	16.810	0.518	2.848	a	a	m	vctk_test_com	13.280	0.413	1.881
13	o	vctk_dev_dif	2.864	0.100	1.134	o	o	f	vctk_test_dif	4.887	0.169	1.495
14	o	vctk_dev_dif	49.970	0.989	166.027	o	a	f	vctk_test_dif	48.050	0.998	146.929
15	o	vctk_dev_dif	26.110	0.760	8.414	a	a	f	vctk_test_dif	31.740	0.847	11.527
	a	vctk_dev_dif	12.410	0.403	2.144	a	a	f	vctk_test_dif	16.200	0.528	3.594
16	o	vctk_dev_dif	1.439	0.052	1.158	o	o	m	vctk_test_dif	2.067	0.072	1.817
17	o	vctk_dev_dif	53.950	1.000	167.511	o	a	m	vctk_test_dif	53.850	1.000	167.824
18	o	vctk_dev_dif	30.920	0.839	23.797	a	a	m	vctk_test_dif	30.940	0.834	23.842
	a	vctk_dev_dif	10.920	0.373	2.210	a	a	m	vctk_test_dif	10.910	0.368	2.176

Table 1: ***Baseline-1***: Speaker verifiability achieved by the ASV_{eval} model trained on the original (**o**) and anonymized (**a**) speech data. The results highlighted in blue are the new results obtained using the ASV_{eval} model trained on anonymized speech data, and the other results in the table correspond to the results reported in [2], Table 8.

#	ASR_{eval}	Dev set	WER, %		Data	Test set	WER, %	
			LM$_s$	LM$_l$			LM$_s$	LM$_l$
1	o	libri_dev	5.24	3.84	o	libri_test	5.55	4.17
2	o	libri_dev	12.19	8.77	a	libri_test	11.77	8.88
	a	libri_dev	5.62	4.11	a	libri_test	5.77	4.34
3	o	vctk_dev	14.00	10.78	o	vctk_test	16.38	12.80
4	o	vctk_dev	30.10	25.56	a	vctk_test	33.25	28.22
	a	vctk_dev	14.83	11.64	a	vctk_test	17.52	13.86

Table 4: ***Baseline-2***: ASR decoding error achieved by ASR_{eval} trained on the original data and anonymized data. The results highlighted in blue are the new results obtained using the ASV_{eval} model trained on anonymized speech data, and the other results in the table correspond to the results reported in [2], Table 11.

These results demonstrate that training ASR_{eval} on anonymized speech data significantly decreases the WER achieved when decoding anonymized speech data in comparison with the case when the model is trained on the original data. For Baseline-1, training ASR_{eval} on anonymized speech data leads to 19–31% relative WER reduction in comparison with training ASR_{eval} on the original data. For Baseline-1, the corresponding relative WER reduction is 47–54%.

Furthermore, training ASV_{eval} on anonymized data rather than original data leads to significant EER reduction for both baseline systems in the case when the enrollment and trial data are anonymized. This reduction is in the order of 40–78% relative for Baseline-1, and 46–90% relative for Baseline-2.

#	ASV_{eval}	**Dev. set**	**EER,%**	$\mathbf{C_{llr}^{min}}$	$\mathbf{C_{llr}}$	**Enroll**	**Trial**	**Gender**	**Test set**	**EER,%**	$\mathbf{C_{llr}^{min}}$	$\mathbf{C_{llr}}$
1	o	libri_dev	8.807	0.305	42.903	o	o	f	libri_test	7.664	0.184	26.808
2	o	libri_dev	35.370	0.821	116.892	o	a	f	libri_test	26.090	0.685	115.571
3	o	libri_dev	23.440	0.621	11.726	a	a	f	libri_test	15.330	0.490	12.553
	a	libri_dev	10.940	0.351	46.154	a	a	f	libri_test	8.029	0.204	30.376
4	o	libri_dev	1.242	0.035	14.294	o	o	m	libri_test	1.114	0.041	15.342
5	o	libri_dev	17.860	0.526	105.715	o	a	m	libri_test	17.820	0.498	106.434
6	o	libri_dev	10.560	0.359	11.951	a	a	m	libri_test	8.241	0.263	15.376
	a	libri_dev	1.087	0.035	16.333	a	a	m	libri_test	1.559	0.045	14.915
7	o	vctk_dev_com	2.616	0.088	0.869	o	o	f	vctk_test_com	2.890	0.092	0.861
8	o	vctk_dev_com	34.300	0.877	85.902	o	a	f	vctk_test_com	30.640	0.807	93.967
9	o	vctk_dev_com	11.630	0.366	43.551	a	a	f	vctk_test_com	14.450	0.465	42.734
	a	vctk_dev_com	4.070	0.143	1.368	a	a	f	vctk_test_com	6.358	0.211	1.435
10	o	vctk_dev_com	1.425	0.050	1.555	o	o	m	vctk_test_com	1.130	0.036	1.042
11	o	vctk_dev_com	23.930	0.669	90.757	o	a	m	vctk_test_com	24.290	0.713	99.336
12	o	vctk_dev_com	10.540	0.316	24.986	a	a	m	vctk_test_com	11.860	0.349	28.225
	a	vctk_dev_com	3.134	0.103	1.412	a	a	m	vctk_test_com	2.542	0.072	1.205
13	o	vctk_dev_dif	2.920	0.101	1.135	o	o	f	vctk_test_dif	4.938	0.169	1.492
14	o	vctk_dev_dif	35.540	0.907	90.540	o	a	f	vctk_test_dif	30.040	0.794	93.211
15	o	vctk_dev_dif	15.830	0.503	39.811	a	a	f	vctk_test_dif	16.920	0.546	41.341
	a	vctk_dev_dif	3.537	0.122	0.798	a	a	f	vctk_test_dif	9.053	0.308	2.823
16	o	vctk_dev_dif	1.439	0.052	1.155	o	o	m	vctk_test_dif	2.067	0.072	1.816
17	o	vctk_dev_dif	28.240	0.741	98.419	o	a	m	vctk_test_dif	28.240	0.720	101.704
18	o	vctk_dev_dif	11.220	0.384	23.093	a	a	m	vctk_test_dif	12.230	0.397	25.064
	a	vctk_dev_dif	3.474	0.127	1.072	a	a	m	vctk_test_dif	4.133	0.147	1.936

Table 3: ***Baseline-2***: Speaker verifiability achieved by the ASV_{eval} model trained on the original (**o**) and anonymized (**a**) speech data. The results highlighted in blue are the new results obtained using the ASV_{eval} model trained on anonymized speech data, and the other results in the table correspond to the results reported in [2], Table 10.

4 Conclusions

On the example of the VoicePrivacy 2020 challenge baseline anonymization systems, we demonstrated that a strong attack model can be developed when an attacker has access to anonymized speech data. This attack model significantly outperforms all the attack models considered in the standard challenge setup. Also we observed that anonymized data can be successfully used for ASR training. In particular, this allows to collect speech data from the users for model training while preserving privacy.

References

[1] N. Tomashenko, B. M. L. Srivastava, X. Wang, E. Vincent, A. Nautsch, J. Yamagishi, N. Evans, J. Patino, J.-F. Bonastre, P.-G. Noé, and M. Todisco, "Introducing the VoicePrivacy Initiative," in *Proc. Interspeech 2020*, 2020, pp. 1693–1697. [Online]. Available: http://dx.doi.org/10.21437/Interspeech.2020-1333

[2] N. Tomashenko, B. M. L. Srivastava, X. Wang, E. Vincent, A. Nautsch, J. Yamagishi, N. Evans, , J. Patino, J.-F. Bonastre, P.-G. Noé, and M. Todisco, "The VoicePrivacy 2020 Challenge evaluation plan," 2020. [Online]. Available: https://www.voiceprivacychallenge.org/docs/VoicePrivacy_2020_Eval_Plan_v1_3.pdf