More results for paper: Detecting Voice Cloning Attacks via Timbre Watermarking

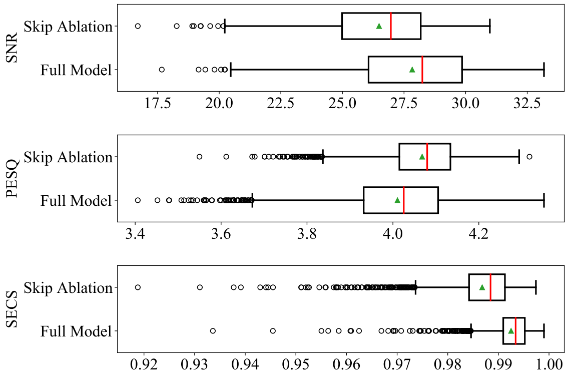

Fig. 1: Fidelity comparison between the Full Model and the model without Skip Concatenation. Green triangles represent the mean values and red lines indicate the median values.

A. *More Ablation Studies*

The Influence of Skip Concatenation. we conduct additional investigations into the Skip Concatenation technique illustrated in framework shown in original paper (Fig. 4) through ablation experiments. Empirically, the Skip Concatenation operation has been found to enhance the model's ability to handle various levels of information, while simultaneously enhancing the stability of the deep model. As demonstrated in Table I, this improvement contributes to an increased level of watermarking robustness. Furthermore, the Skip Concatenation operation serves to mitigate information loss. As depicted in Fig. 1, the results show the improved fidelity compared to the ablation model.

B. *Integrity Verification*

Here, we assume that the speech without a watermark may be pulled from a watermark-free domain to another domain after various processing operations. For integrity, we do not want these processed speeches to fall into the watermarked domain. In order to verify that the scheme does not incorrectly verify the existence of a watermark on the watermark-free speech, we try to apply various pre-processing on the watermark-free speech and further perform watermark extraction to verify the accuracy of watermark extraction. As shown in Table II, the various processing operations do not make the unwatermarked speech be pulled into the space containing the watermark (the extracted information is a random sequence).

TABLE I: The impact of different postprocessing operations on the speech quality and robustness of the watermarking model without Skip Concatenation. ACC* represents the extraction accuracy of the Full Model.

Preprocessing	Parameter	Quality			ACC↑	ACC-ACC*
		SNR↑	PESQ↑	SECS↑		
Resampling	16 kHz	37.5646	4.4991	1.0000	1.0000	0.0000
	8 kHz	17.4609	4.4985	0.9201	0.9856	-0.0084
Amplitude Scaling	20%	1.9382	4.4907	0.9603	1.0000	0.0000
	40%	4.4368	4.4970	0.9622	1.0000	0.0000
	60%	7.9589	4.4984	0.9785	1.0000	0.0000
	80%	13.9789	4.4989	0.9944	1.0000	0.0000
MP3 Compression	8 kbps	9.1037	2.1904	0.8000	0.9473	0.0287
	16 kbps	13.4830	3.3000	0.9703	0.9966	-0.0026
	24 kbps	15.6557	3.8254	0.9909	0.9999	0.0000
	32 kbps	17.7468	3.9695	0.9968	1.0000	0.0000
	40 kbps	19.2804	4.0998	0.9979	1.0000	0.0000
	48 kbps	21.2728	4.2509	0.9988	1.0000	0.0000
	56 kbps	23.1696	4.3433	0.9992	1.0000	0.0000
	64 kbps	24.2805	4.3888	0.9993	1.0000	0.0000
Recount	8 bps	22.9386	3.0350	0.9724	0.9941	-0.0054
Median Filtering	5 Samples	15.0288	3.5303	0.9417	0.9997	-0.0003
	15 Samples	8.8760	2.3857	0.7769	0.9947	0.0014
	25 Samples	5.3526	1.9653	0.7251	0.9836	0.0030
	35 Samples	3.2232	1.6844	0.6791	0.9484	0.0082
Low Pass Filtering	2000 Hz	12.9725	3.8897	0.7531	0.9283	0.0253
High Pass Filtering	500 Hz	3.7764	3.7998	0.6620	1.0000	0.0000
Gaussian Noise	20 dB	20.0001	2.8979	0.8920	0.9689	-0.0273
	25 dB	24.9990	3.2840	0.9600	0.9934	-0.0061
	30 dB	29.9972	3.6664	0.9909	0.9994	-0.0006
	35 dB	34.9924	4.0039	0.9982	0.9997	-0.0003
	40 dB	39.9852	4.2502	0.9995	1.0000	0.0000

C. *Combing Multiple Attack Strategies*

See Table III ∼ Table XIII. We further consider combining different attack strategies to destroy the proposed method. This entails the integration of diverse attack schemes, encompassing regular preprocessing (Table III, Table IV and Table V), harmful preprocessing (Table VI, Table VII and Table VIII), domain-adversarial training (Table X), VAE reconstruction (Table IX) and watermark overwriting (Table XI, Table XII and Table XIII). In a ntushell, more severe attack strategies will further destroy the utility of voice cloning, while the proposed method is still somewhat effective. For example, taking resampling 16 KHZ as pre-processing and MP3 compression 16Kbps as post-processing, compared with only pre-processing, ACC suffers a slight degradation (ACC:100% → 99.94%) but the quality degrades by a large margin (SECS: 1.000 → 0.8575).

References

[1] Zhenghui Liu, Yuankun Huang, and Jiwu Huang. Patchwork-based audio watermarking robust against de-synchronization and recapturing attacks. *IEEE transactions on information forensics and security*, 14(5):1171–1180, 2018.

[2] Juan Zhao, Tianrui Zong, Yong Xiang, Longxiang Gao, Wanlei Zhou, and Gleb Beliakov. Desynchronization attacks resilient watermarking method based on frequency singular value coefficient modification. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, 29:2282–2295, 2021.

TABLE II: The impact of different preprocessing on speech quality and wm-free speech's watermark extraction.

Preprocessing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resample	16 kHz	35.6733	4.4996	1.0000	0.4991
	8 kHz	17.0287	4.4990	0.9069	0.5052
Amplitude Scaling	20%	1.9382	4.4922	0.9582	0.5002
	40%	4.4368	4.4975	0.9603	0.4999
	60%	7.9589	4.4988	0.9776	0.5008
	80%	13.9790	4.4991	0.9943	0.5005
MP3 Compression	8 kbps	8.9250	2.1765	0.7714	0.5030
	16 kbps	13.0672	3.3063	0.9606	0.4956
	24 kbps	15.2163	3.8745	0.9895	0.5008
	32 kbps	17.2447	4.0198	0.9964	0.5005
	40 kbps	18.7567	4.1426	0.9976	0.5038
	48 kbps	20.6943	4.2759	0.9986	0.4992
	56 kbps	22.6779	4.3558	0.9991	0.4987
	64 kbps	23.9003	4.3981	0.9992	0.5015
Recount	8 bps	22.9106	3.1203	0.9738	0.5008
Median Filtering	5 Samples	14.7018	3.5970	0.9435	0.4971
	15 Samples	8.7834	2.4883	0.7808	0.4985
	25 Samples	5.3345	2.0663	0.7286	0.5000
	35 Samples	3.2235	1.7943	0.6835	0.5006
Low Pass Filtering	2000 Hz	12.6986	3.8897	0.7328	0.5027
High Pass Filtering	500 Hz	3.7776	3.8107	0.6584	0.5031
Gaussian Noise	20 dB	20.0002	2.9945	0.9032	0.4969
	25 dB	24.9987	3.3807	0.9644	0.5010
	30 dB	29.9976	3.7492	0.9915	0.4996
	35 dB	34.9929	4.0593	0.9982	0.4992
	40 dB	39.9868	4.2783	0.9994	0.4996
Mel Masking	position-1	1.1221	3.0576	0.5275	0.4972
	position-2	11.595	3.5904	0.8538	0.5031
	position-3	17.1566	3.8715	0.8471	0.5058
	position-4	19.8721	4.306	0.9382	0.5024
	position-5	21.9348	4.4992	0.985	0.5037
	position-6	25.2127	4.4993	0.9962	0.4983
	position-7	30.2568	4.4994	0.9993	0.5016
	position-8	37.8566	4.4996	1	0.4962
	position-9	47.4514	4.4998	1	0.5054
	position-10	67.9461	4.5	1	0.5010

TABLE III: The impact on the speech quality and robustness under adaptive attacks combined by Resampling 16K preprocessing and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.6970	1.0763	0.9121	1.0000
	8 kHz	-2.5804	1.0755	0.7931	1.0000
Amplitude Scaling	20%	-0.1514	1.0633	0.8661	1.0000
	40%	-0.5723	1.0727	0.8664	1.0000
	60%	-1.1886	1.0761	0.8854	1.0000
	80%	-1.9208	1.0752	0.9061	1.0000
MP3 Compression	8 kbps	-2.3744	0.9379	0.6661	0.8876
	16 kbps	-2.4251	1.0389	0.8575	0.9994
	24 kbps	-2.4429	1.0716	0.8862	1.0000
	32 kbps	-2.5187	1.0762	0.9083	1.0000
	40 kbps	-2.5138	1.0628	0.9096	1.0000
	48 kbps	-2.5092	1.0660	0.9105	1.0000
	56 kbps	-2.5073	1.0631	0.9110	1.0000
	64 kbps	-2.5068	1.0637	0.9112	1.0000
Recount	8 bps	-2.5671	1.0208	0.8880	0.9998
Median Filtering	5 Samples	-2.5422	1.0443	0.8313	1.0000
	15 Samples	-2.0577	1.0380	0.7255	0.9948
	25 Samples	-1.3706	0.9776	0.6754	0.9748
	35 Samples	-0.7732	0.8825	0.5926	0.8948
Low Pass Filtering	2000 Hz	-2.5236	1.1163	0.6542	0.8286
High Pass Filtering	500 Hz	-1.5068	1.3068	0.6444	1.0000
Gaussian Noise	20 dB	-2.7257	1.0475	0.8221	0.9984
	25 dB	-2.7122	1.0359	0.8820	1.0000
	30 dB	-2.7079	1.0464	0.9054	1.0000
	35 dB	-2.7066	1.0640	0.9107	1.0000
	40 dB	-2.7061	1.0650	0.9119	1.0000

TABLE IV: The impact on the speech quality and robustness under adaptive attacks combined by Mp3 Compression 64kbps preprocessing and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.4774	1.0363	0.9073	1.0000
	8 kHz	-2.3602	1.0349	0.7986	1.0000
Amplitude Scaling	20%	-0.1418	1.0376	0.8634	1.0000
	40%	-0.5418	1.0338	0.8634	1.0000
	60%	-1.1324	1.0389	0.8765	1.0000
	80%	-1.8389	1.0331	0.8992	1.0000
MP3 Compression	8 kbps	-2.1655	0.9020	0.6511	0.8854
	16 kbps	-2.2128	1.0198	0.8493	0.9990
	24 kbps	-2.2317	1.0485	0.8826	1.0000
	32 kbps	-2.3891	1.0409	0.9028	1.0000
	40 kbps	-2.4262	1.0574	0.9032	1.0000
	48 kbps	-2.4192	1.0370	0.9045	1.0000
	56 kbps	-2.4126	1.0416	0.9052	1.0000
	64 kbps	-2.4099	1.0408	0.9059	1.0000
Recount	8 bps	-2.4609	0.9887	0.8842	0.9996
Median Filtering	5 Samples	-2.3429	1.0004	0.8104	1.0000
	15 Samples	-1.9205	0.9989	0.7162	0.9942
	25 Samples	-1.2878	0.9218	0.6794	0.9654
	35 Samples	-0.7014	0.8368	0.5951	0.8754
Low Pass Filtering	2000 Hz	-2.3137	1.0789	0.6510	0.8406
High Pass Filtering	500 Hz	-1.6076	1.2830	0.6424	1.0000
Gaussian Noise	20 dB	-2.6202	1.0049	0.8087	0.9970
	25 dB	-2.6071	1.0203	0.8735	0.9998
	30 dB	-2.6029	1.0153	0.9004	1.0000
	35 dB	-2.6016	1.0320	0.9063	1.0000
	40 dB	-2.6012	1.0347	0.9074	1.0000

TABLE V: The impact on the speech quality and robustness under adaptive attacks combined by Regular preprocessing (refer to TABLE IV in original paper) and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.5611	1.0730	0.9063	1.0000
	8 kHz	-2.4416	1.0746	0.7948	1.0000
Amplitude Scaling	20%	-0.1451	1.0790	0.8563	1.0000
	40%	-0.5488	1.0791	0.8567	1.0000
	60%	-1.1416	1.0668	0.8729	1.0000
	80%	-1.8485	1.0765	0.8974	1.0000
MP3 Compression	8 kbps	-2.2454	0.9148	0.6512	0.8840
	16 kbps	-2.2935	1.0483	0.8479	0.9988
	24 kbps	-2.3066	1.0764	0.8815	1.0000
	32 kbps	-2.4180	1.0660	0.9020	1.0000
	40 kbps	-2.4283	1.0715	0.9038	1.0000
	48 kbps	-2.4229	1.0801	0.9049	1.0000
	56 kbps	-2.4186	1.0734	0.9051	1.0000
	64 kbps	-2.4174	1.0800	0.9052	1.0000
Recount	8 bps	-2.4683	0.9844	0.8887	0.9998
Median Filtering	5 Samples	-2.4132	1.0571	0.8242	1.0000
	15 Samples	-1.9680	1.0416	0.7225	0.9874
	25 Samples	-1.3558	0.9969	0.6757	0.9556
	35 Samples	-0.7719	0.8823	0.5856	0.8738
Low Pass Filtering	2000 Hz	-2.3910	1.1251	0.6444	0.7664
High Pass Filtering	500 Hz	-1.4417	1.2626	0.6488	1.0000
Gaussian Noise	20 dB	-2.6289	1.0422	0.8179	0.9968
	25 dB	-2.6157	1.0280	0.8781	0.9998
	30 dB	-2.6115	1.0334	0.9011	1.0000
	35 dB	-2.6102	1.0643	0.9057	1.0000
	40 dB	-2.6098	1.0609	0.9062	1.0000

TABLE VI: The impact on the speech quality and robustness under adaptive attacks combined by Mp3 Compression 8kbps preprocessing and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-1.4826	0.8243	0.6675	0.8874
	8 kHz	-1.4828	0.8369	0.6675	0.8822
Amplitude Scaling	20%	-0.0705	0.8473	0.6595	0.8968
	40%	-0.2761	0.8347	0.6595	0.8962
	60%	-0.5974	0.8326	0.6596	0.8968
	80%	-1.0084	0.8305	0.6632	0.8978
MP3 Compression	8 kbps	-1.3637	0.8100	0.6511	0.8706
	16 kbps	-1.3591	0.8439	0.6648	0.8852
	24 kbps	-1.3588	0.8493	0.6662	0.8824
	32 kbps	-1.3586	0.8363	0.6664	0.8854
	40 kbps	-1.3586	0.8528	0.6664	0.8914
	48 kbps	-1.3585	0.8439	0.6664	0.8970
	56 kbps	-1.3585	0.8477	0.6664	0.8992
	64 kbps	-1.3585	0.8360	0.6664	0.8956
Recount	8 bps	-1.3651	0.8824	0.6679	0.8322
Median Filtering	5 Samples	-1.4562	0.9410	0.6767	0.8938
	15 Samples	-1.1424	1.0001	0.6979	0.8914
	25 Samples	-0.7597	0.9078	0.6564	0.8314
	35 Samples	-0.4462	0.7998	0.5599	0.8368
Low Pass Filtering	2000 Hz	-1.4588	0.9530	0.6416	0.8560
High Pass Filtering	500 Hz	-0.7249	1.0194	0.5216	0.9266
Gaussian Noise	20 dB	-1.4950	0.9527	0.6514	0.8724
	25 dB	-1.4865	0.9424	0.6684	0.8788
	30 dB	-1.4838	0.9212	0.6723	0.8866
	35 dB	-1.4830	0.9056	0.6697	0.8912
	40 dB	-1.4827	0.9021	0.6682	0.8908

TABLE VII: The impact on the speech quality and robustness under adaptive attacks combined by Low Pass Filtering 2000 Hz preprocessing and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.5652	1.0766	0.6481	0.9492
	8 kHz	-2.5655	1.0835	0.6481	0.9554
Amplitude Scaling	20%	-0.1413	1.0893	0.6445	0.9750
	40%	-0.5352	1.0845	0.6445	0.9734
	60%	-1.1163	1.0843	0.6464	0.9668
	80%	-1.8126	1.0888	0.6473	0.9570
MP3 Compression	8 kbps	-2.3780	1.0161	0.6471	0.8676
	16 kbps	-2.3746	1.0814	0.6459	0.9506
	24 kbps	-2.3735	1.0878	0.6472	0.9668
	32 kbps	-2.3734	1.0790	0.6474	0.9594
	40 kbps	-2.3733	1.0809	0.6475	0.9628
	48 kbps	-2.3733	1.0821	0.6475	0.9646
	56 kbps	-2.3733	1.0934	0.6475	0.9686
	64 kbps	-2.3733	1.0887	0.6475	0.9724
Recount	8 bps	-2.4338	0.9672	0.6496	0.8638
Median Filtering	5 Samples	-2.5416	1.1020	0.6538	0.9564
	15 Samples	-2.1369	1.0153	0.6833	0.8746
	25 Samples	-1.4919	0.9339	0.6524	0.8006
	35 Samples	-0.8831	0.8446	0.5870	0.8020
Low Pass Filtering	2000 Hz	-2.5583	1.0873	0.6516	0.9832
High Pass Filtering	500 Hz	-1.2457	1.3786	0.4906	0.8890
Gaussian Noise	20 dB	-2.5841	1.0019	0.6530	0.8968
	25 dB	-2.5712	1.0145	0.6576	0.9154
	30 dB	-2.5672	1.0295	0.6551	0.9292
	35 dB	-2.5658	1.0524	0.6517	0.9404
	40 dB	-2.5654	1.0687	0.6494	0.9428

TABLE VIII: The impact on the speech quality and robustness under adaptive attacks combined by Harmful preprocessing (refer to TABLE IV in original paper) and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.3467	1.0331	0.6567	0.8956
	8 kHz	-2.3469	1.0291	0.6567	0.8956
Amplitude Scaling	20%	-0.1256	1.0332	0.6508	0.9402
	40%	-0.4795	1.0296	0.6508	0.9294
	60%	-1.0072	1.0290	0.6524	0.9226
	80%	-1.6471	1.0310	0.6542	0.9176
MP3 Compression	8 kbps	-2.1717	0.9753	0.6468	0.8872
	16 kbps	-2.1681	1.0261	0.6538	0.9458
	24 kbps	-2.1678	1.0304	0.6555	0.9166
	32 kbps	-2.1676	1.0285	0.6558	0.8974
	40 kbps	-2.1676	1.0251	0.6558	0.9124
	48 kbps	-2.1675	1.0166	0.6558	0.9260
	56 kbps	-2.1675	1.0305	0.6558	0.9338
	64 kbps	-2.1676	1.0384	0.6558	0.9376
Recount	8 bps	-2.2161	0.9630	0.6585	0.8534
Median Filtering	5 Samples	-2.3231	1.0792	0.6645	0.9528
	15 Samples	-1.9520	1.0472	0.6941	0.9166
	25 Samples	-1.3512	0.9786	0.6588	0.8212
	35 Samples	-0.7601	0.8782	0.5795	0.8150
Low Pass Filtering	2000 Hz	-2.3349	1.0805	0.6494	0.9320
High Pass Filtering	500 Hz	-1.0542	1.2453	0.5039	0.8916
Gaussian Noise	20 dB	-2.3644	1.0013	0.6615	0.8984
	25 dB	-2.3523	1.0102	0.6631	0.9188
	30 dB	-2.3485	1.0155	0.6620	0.9368
	35 dB	-2.3473	1.0158	0.6594	0.9440
	40 dB	-2.3469	1.0421	0.6576	0.9444

TABLE IX: The impact on the speech quality and robustness under adaptive attacks combined by VAE Reconstruction preprocessing and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.1132	1.0064	0.9006	1.0000
	8 kHz	-2.0458	1.0085	0.7849	1.0000
Amplitude Scaling	20%	-0.1113	1.0013	0.8667	1.0000
	40%	-0.4253	1.0001	0.8667	1.0000
	60%	-0.8973	1.0034	0.8714	1.0000
	80%	-1.4754	1.0029	0.8902	1.0000
MP3 Compression	8 kbps	-1.8712	0.8344	0.6519	0.8488
	16 kbps	-1.9150	0.9689	0.8344	0.9956
	24 kbps	-1.9294	0.9969	0.8723	0.9984
	32 kbps	-1.9607	1.0067	0.8947	0.9966
	40 kbps	-1.9561	1.0037	0.8957	0.9988
	48 kbps	-1.9520	1.0022	0.8973	0.9992
	56 kbps	-1.9508	1.0016	0.8983	0.9996
	64 kbps	-1.9504	1.0031	0.8988	1.0000
Recount	8 bps	-1.9775	0.9173	0.8801	0.9948
Median Filtering	5 Samples	-2.0041	0.9864	0.8302	0.9994
	15 Samples	-1.6191	0.9706	0.7103	0.9654
	25 Samples	-1.1075	0.8893	0.6562	0.8972
	35 Samples	-0.6462	0.7831	0.5717	0.7456
Low Pass Filtering	2000 Hz	-1.9955	1.0428	0.6454	0.6286
High Pass Filtering	500 Hz	-1.1293	1.2265	0.6230	1.0000
Gaussian Noise	20 dB	-2.1307	0.9620	0.8141	0.9878
	25 dB	-2.1197	0.9755	0.8732	0.9966
	30 dB	-2.1162	0.9810	0.8945	0.9992
	35 dB	-2.1151	0.9985	0.8994	0.9996
	40 dB	-2.1147	0.9982	0.9005	1.0000

TABLE X: The impact on the speech quality and robustness under adaptive attacks combined by Domain-adversarial training as preprocessing and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.8678	0.8798	0.8842	1.0000
	8 kHz	-2.4934	0.8823	0.7569	1.0000
Amplitude Scaling	20%	-0.2059	0.8823	0.8372	1.0000
	40%	-0.7636	0.8758	0.8373	1.0000
	60%	-1.5474	0.8748	0.8566	1.0000
	80%	-2.4403	0.8735	0.8775	1.0000
MP3 Compression	8 kbps	-2.2868	0.7229	0.6483	0.8942
	16 kbps	-2.3555	0.8473	0.8249	1.0000
	24 kbps	-2.3758	0.8860	0.8535	1.0000
	32 kbps	-2.9885	0.8731	0.8776	1.0000
	40 kbps	-3.1691	0.8740	0.8791	1.0000
	48 kbps	-3.1554	0.8833	0.8808	1.0000
	56 kbps	-3.1409	0.8657	0.8823	1.0000
	64 kbps	-3.1339	0.8757	0.8828	1.0000
Recount	8 bps	-3.2272	0.8045	0.8673	0.9996
Median Filtering	5 Samples	-2.5592	0.8361	0.7812	1.0000
	15 Samples	-2.0092	0.8876	0.6885	0.9984
	25 Samples	-1.3483	0.8645	0.6520	0.9770
	35 Samples	-0.7814	0.7448	0.5530	0.8800
Low Pass Filtering	2000 Hz	-2.4262	0.9693	0.6296	0.8018
High Pass Filtering	500 Hz	-2.4610	1.0999	0.6393	1.0000
Gaussian Noise	20 dB	-3.3859	0.8990	0.7799	0.9978
	25 dB	-3.3704	0.8851	0.8421	0.9994
	30 dB	-3.3655	0.8928	0.8731	0.9998
	35 dB	-3.3639	0.8790	0.8820	1.0000
	40 dB	-3.3634	0.8893	0.8837	1.0000

TABLE XI: The impact on the speech quality and robustness under adaptive attacks combined by the watermark overwriting attacks (adopt FSVC [2]) and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.6914	1.0390	0.9113	1.0000
	8 kHz	-2.5565	1.0376	0.7906	1.0000
Amplitude Scaling	20%	-0.1580	1.0358	0.8658	1.0000
	40%	-0.5987	1.0419	0.8661	1.0000
	60%	-1.2416	1.0405	0.8839	1.0000
	80%	-2.0013	1.0314	0.9058	1.0000
MP3 Compression	8 kbps	-2.3511	0.8877	0.6549	0.8830
	16 kbps	-2.3982	1.0132	0.8562	0.9994
	24 kbps	-2.4121	1.0247	0.8840	1.0000
	32 kbps	-2.5914	1.0268	0.9076	1.0000
	40 kbps	-2.6216	1.0293	0.9085	1.0000
	48 kbps	-2.6138	1.0336	0.9094	1.0000
	56 kbps	-2.6083	1.0305	0.9102	1.0000
	64 kbps	-2.6050	1.0281	0.9106	1.0000
Recount	8 bps	-2.6685	0.9724	0.8888	0.9998
Median Filtering	5 Samples	-2.5397	0.9926	0.8119	1.0000
	15 Samples	-2.1238	0.9937	0.7092	0.9970
	25 Samples	-1.4311	0.9229	0.6788	0.9750
	35 Samples	-0.8009	0.8241	0.5875	0.8838
Low Pass Filtering	2000 Hz	-2.5059	1.0920	0.6499	0.8014
High Pass Filtering	500 Hz	-1.6218	1.2533	0.6402	1.0000
Gaussian Noise	20 dB	-2.8317	1.0049	0.8100	0.9990
	25 dB	-2.8178	1.0132	0.8763	0.9998
	30 dB	-2.8134	1.0138	0.9036	1.0000
	35 dB	-2.8121	1.0201	0.9100	1.0000
	40 dB	-2.8116	1.0292	0.9111	1.0000

TABLE XII: The impact on the speech quality and robustness under adaptive attacks combined by the watermark overwriting attacks (adopt Patchwork method [1]) and different postprocessing operations below.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-2.5780	1.0762	0.9101	1.0000
	8 kHz	-2.4405	1.0764	0.7854	1.0000
Amplitude Scaling	20%	-0.1492	1.0687	0.8629	1.0000
	40%	-0.5648	1.0765	0.8630	1.0000
	60%	-1.1751	1.0716	0.8786	1.0000
	80%	-1.9020	1.0737	0.9031	1.0000
MP3 Compression	8 kbps	-2.2412	0.8936	0.6516	0.8830
	16 kbps	-2.2935	1.0233	0.8549	0.9998
	24 kbps	-2.3080	1.0620	0.8859	1.0000
	32 kbps	-2.4686	1.0509	0.9073	1.0000
	40 kbps	-2.4998	1.0622	0.9080	1.0000
	48 kbps	-2.4927	1.0605	0.9085	1.0000
	56 kbps	-2.4864	1.0666	0.9091	1.0000
	64 kbps	-2.4836	1.0682	0.9092	1.0000
Recount	8 bps	-2.5407	1.0009	0.8885	1.0000
Median Filtering	5 Samples	-2.4171	1.0335	0.8101	1.0000
	15 Samples	-1.9733	1.0130	0.7121	0.9950
	25 Samples	-1.3396	0.9488	0.6716	0.9664
	35 Samples	-0.7637	0.8551	0.5837	0.8760
Low Pass Filtering	2000 Hz	-2.3838	1.1149	0.6519	0.8062
High Pass Filtering	500 Hz	-1.5573	1.2648	0.6426	1.0000
Gaussian Noise	20 dB	-2.7024	1.0381	0.8162	0.9980
	25 dB	-2.6891	1.0279	0.8760	0.9998
	30 dB	-2.6848	1.0413	0.9027	1.0000
	35 dB	-2.6834	1.0369	0.9089	1.0000
	40 dB	-2.6830	1.0644	0.9099	1.0000

TABLE XIII: The impact on the speech quality and robustness under adaptive attacks combined by the watermark overwriting attacks (adopt the proposed method *) and different postprocessing operations below. * indicates that the attacker trains his own embedding and extraction models guided by the proposed method.

Processing	Parameter	Quality			ACC↑
		SNR↑	PESQ↑	SECS↑	
Resampling	16 kHz	-1.5347	1.0018	0.8789	0.9352
	8 kHz	-1.4347	0.9887	0.7610	0.8970
Amplitude Scaling	20%	-0.0786	0.9791	0.8509	0.9860
	40%	-0.3062	0.9958	0.8509	0.9826
	60%	-0.6591	0.9837	0.8509	0.9678
	80%	-1.1067	0.9989	0.8603	0.9506
MP3 Compression	8 kbps	-1.3070	0.8800	0.6236	0.7148
	16 kbps	-1.3382	0.9774	0.8214	0.8456
	24 kbps	-1.3424	0.9968	0.8508	0.9660
	32 kbps	-1.4736	1.0054	0.8759	0.9706
	40 kbps	-1.4904	1.0067	0.8769	0.9736
	48 kbps	-1.4876	1.0024	0.8770	0.9804
	56 kbps	-1.4856	0.9922	0.8768	0.9782
	64 kbps	-1.4844	0.9972	0.8764	0.9812
Recount	8 bps	-1.4898	0.8892	0.8617	0.8598
Median Filtering	5 Samples	-1.4270	0.9382	0.7946	0.9414
	15 Samples	-1.1317	0.9416	0.7159	0.8780
	25 Samples	-0.7644	0.9025	0.6769	0.8146
	35 Samples	-0.4441	0.7888	0.5955	0.7442
Low Pass Filtering	2000 Hz	-1.4013	1.0483	0.6257	0.6554
High Pass Filtering	500 Hz	-0.8945	1.2185	0.6348	0.9314
Gaussian Noise	20 dB	-1.6317	0.9561	0.8136	0.8564
	25 dB	-1.6227	0.9734	0.8647	0.8816
	30 dB	-1.6198	0.9680	0.8798	0.9334
	35 dB	-1.6189	0.9869	0.8800	0.9676
	40 dB	-1.6186	0.9821	0.8793	0.9766