Appendices

Additional material:

- Appendix A: the 1,000 clean chosen audio samples (100 clean audio samples per command word × 10 command words).
- Appendix B: Grid search leading to Speech-EA parameter values.
- Appendix C: Representative Mel spectrograms for each command word assessed on a clean and corresponding adversarial audio.

Appendix A: Chosen Clean Audio Samples

Table 3 lists the 100 clean audio file samples for each command taken from the Synthetic Speech Commands Dataset (referred to in the Experimental Settings Section of the article). They are selected because their transcriptions by Wav2 Vec 2.0 are accurate. The first number 1029 in the column `DOWN` and the raw $\mathcal{A}_1$ means that $\mathcal{A}_1(\texttt{DOWN}) = 1029.\texttt{wav}$ is the audio file we selected in the Synthetic Speech Commands Dataset. This audio file, where the word `DOWN` is spoken, is clean since its transcription by Wav2Vec 2.0 is correct. *Mutatis mutandis* for the other numbers per command for each command.

Audio file	DOWN	GO	LEFT	NO	OFF	ON	RIGHT	STOP	UP	YES
$\mathcal{A}_1$	1029	167	102	108	107	1487	104	102	103	112
$\mathcal{A}_2$	1033	173	103	112	119	1488	118	103	108	113
$\mathcal{A}_3$	1378	178	107	114	127	1489	127	104	112	114
$\mathcal{A}_4$	1383	182	108	12	134	1492	132	174	12	117
$\mathcal{A}_5$	1384	184	113	14	136	1493	137	179	13	118
$\mathcal{A}_6$	1392	189	127	153	137	1497	139	203	14	119
$\mathcal{A}_7$	1393	19	129	167	146	1506	147	204	153	12
$\mathcal{A}_8$	1394	194	13	168	148	1508	148	208	169	122
$\mathcal{A}_9$	1398	209	133	169	153	1532	152	213	18	123
$\mathcal{A}_{10}$	1399	237	137	17	156	1543	153	214	183	124
$\mathcal{A}_{11}$	1402	249	139	173	192	1544	158	218	184	127
$\mathcal{A}_{12}$	1408	252	14	174	196	1548	159	228	194	128
$\mathcal{A}_{13}$	1409	254	142	18	197	1552	163	237	198	129
$\mathcal{A}_{14}$	1412	263	143	182	198	1554	164	238	199	13
$\mathcal{A}_{15}$	1413	264	144	184	202	1559	173	247	2	132
$\mathcal{A}_{16}$	1414	267	147	187	207	1569	184	249	202	133
$\mathcal{A}_{17}$	1417	269	153	189	211	1578	219	252	204	134
$\mathcal{A}_{18}$	1418	27	158	192	212	1584	222	257	212	137
$\mathcal{A}_{19}$	1419	272	163	194	214	1604	223	258	214	138
$\mathcal{A}_{20}$	1422	274	164	198	22	1609	224	259	217	139
$\mathcal{A}_{21}$	1423	279	168	2	222	1612	228	262	218	14

Audio file	DOWN	GO	LEFT	NO	OFF	ON	RIGHT	STOP	UP	YES
$\mathcal{A}_{22}$	1427	3	172	202	232	1614	229	268	219	142
$\mathcal{A}_{23}$	1437	302	173	204	238	1619	234	269	264	143
$\mathcal{A}_{24}$	1439	303	177	207	239	1627	237	272	29	144
$\mathcal{A}_{25}$	1443	307	178	208	249	1639	239	273	303	147
$\mathcal{A}_{26}$	1444	309	182	209	251	1643	242	274	304	148
$\mathcal{A}_{27}$	1447	313	184	212	258	1651	244	28	308	149
$\mathcal{A}_{28}$	1453	314	187	213	261	1652	247	282	309	152
$\mathcal{A}_{29}$	1454	317	193	214	263	1659	248	283	314	153
$\mathcal{A}_{30}$	1457	318	197	219	27	1662	252	287	317	154
$\mathcal{A}_{31}$	1458	323	198	222	277	1664	254	29	319	157
$\mathcal{A}_{32}$	1462	324	2	223	29	1666	258	294	32	158
$\mathcal{A}_{33}$	1467	327	203	224	339	1667	262	298	323	162
$\mathcal{A}_{34}$	1468	328	204	229	343	1669	263	302	329	163
$\mathcal{A}_{35}$	1472	329	208	232	344	1677	264	307	33	164
$\mathcal{A}_{36}$	1477	33	209	234	348	1693	272	309	34	167
$\mathcal{A}_{37}$	1482	34	212	237	349	1706	273	312	39	168
$\mathcal{A}_{38}$	208	38	213	239	354	1708	3	314	4	169
$\mathcal{A}_{39}$	39	39	214	24	357	1714	334	317	426	17
$\mathcal{A}_{40}$	394	4	217	244	359	1719	384	318	43	172
$\mathcal{A}_{41}$	397	406	218	247	363	1748	388	322	44	173
$\mathcal{A}_{42}$	398	421	22	252	368	1783	393	323	446	174
$\mathcal{A}_{43}$	399	423	23	257	369	1824	404	324	453	177
$\mathcal{A}_{44}$	402	424	243	258	371	1828	409	328	456	179
$\mathcal{A}_{45}$	404	426	257	259	374	1829	424	329	461	18
$\mathcal{A}_{46}$	407	429	264	262	379	1831	433	33	466	182
$\mathcal{A}_{47}$	408	44	268	263	381	1832	434	342	467	183
$\mathcal{A}_{48}$	413	441	269	264	389	1833	437	349	468	184
$\mathcal{A}_{49}$	417	442	273	267	39	1834	442	354	469	187
$\mathcal{A}_{50}$	419	443	274	268	401	1877	443	358	473	188
$\mathcal{A}_{51}$	422	444	278	27	402	1882	447	359	474	189
$\mathcal{A}_{52}$	423	446	279	279	411	1889	448	362	476	19
$\mathcal{A}_{53}$	428	448	287	28	412	1894	449	363	477	192
$\mathcal{A}_{54}$	429	449	3	282	413	1896	461	364	478	193
$\mathcal{A}_{55}$	433	453	308	287	42	1897	463	369	479	194
$\mathcal{A}_{56}$	437	457	319	29	421	1903	464	372	48	197
$\mathcal{A}_{57}$	439	459	32	292	422	1906	467	373	481	198
$\mathcal{A}_{58}$	442	466	324	297	423	1907	469	377	484	199
$\mathcal{A}_{59}$	444	468	329	3	427	1908	473	379	487	203
$\mathcal{A}_{60}$	447	469	33	302	429	1909	478	382	488	204
$\mathcal{A}_{61}$	457	471	359	307	431	1911	484	42	489	207
$\mathcal{A}_{62}$	469	473	369	308	432	1916	486	43	491	208
$\mathcal{A}_{63}$	473	474	373	309	436	1918	488	442	493	209
$\mathcal{A}_{64}$	477	48	387	312	437	1921	489	447	496	212
$\mathcal{A}_{65}$	482	482	388	313	439	1923	498	448	498	213

Audio file	DOWN	GO	LEFT	NO	OFF	ON	RIGHT	STOP	UP	YES
$\mathcal{A}_{66}$	483	483	389	314	446	1924	52	452	503	214
$\mathcal{A}_{67}$	484	484	39	317	447	1926	534	453	504	217
$\mathcal{A}_{68}$	487	487	392	318	451	1928	539	454	507	218
$\mathcal{A}_{69}$	493	488	393	319	452	1929	54	457	512	22
$\mathcal{A}_{70}$	534	489	394	32	456	1932	542	458	514	222
$\mathcal{A}_{71}$	538	492	397	324	457	1933	544	462	52	224
$\mathcal{A}_{72}$	884	499	4	328	458	1934	552	463	524	228
$\mathcal{A}_{73}$	892	504	43	329	459	1936	553	464	53	23
$\mathcal{A}_{74}$	893	524	44	33	462	1941	559	467	539	233
$\mathcal{A}_{75}$	898	528	47	332	464	1943	567	469	542	24
$\mathcal{A}_{76}$	903	53	49	333	468	1944	612	473	543	27
$\mathcal{A}_{77}$	907	669	52	34	469	1946	614	477	549	28
$\mathcal{A}_{78}$	908	673	53	357	473	1947	618	478	569	29
$\mathcal{A}_{79}$	909	674	54	358	476	1948	623	479	594	3
$\mathcal{A}_{80}$	912	684	57	37	477	1952	624	48	648	32
$\mathcal{A}_{81}$	913	689	58	39	478	1953	629	482	652	33
$\mathcal{A}_{82}$	914	734	59	4	479	1956	634	487	653	34
$\mathcal{A}_{83}$	917	739	63	43	48	1958	642	488	668	37
$\mathcal{A}_{84}$	918	749	64	44	481	1959	643	49	677	38
$\mathcal{A}_{85}$	922	754	67	47	482	1962	649	492	694	39
$\mathcal{A}_{86}$	923	763	69	49	483	1963	653	493	709	4
$\mathcal{A}_{87}$	924	769	73	52	486	1966	657	504	713	42
$\mathcal{A}_{88}$	927	798	74	53	487	1967	658	522	714	43
$\mathcal{A}_{89}$	928	799	77	54	488	1968	702	523	719	44
$\mathcal{A}_{90}$	929	8	78	63	49	1972	718	532	773	47
$\mathcal{A}_{91}$	933	804	8	7	491	1973	72	538	774	48
$\mathcal{A}_{92}$	937	807	83	72	492	1977	723	539	786	49
$\mathcal{A}_{93}$	963	808	84	77	493	1978	724	54	789	53
$\mathcal{A}_{94}$	967	812	88	79	494	1979	739	63	794	54
$\mathcal{A}_{95}$	968	813	89	8	497	1981	743	69	798	7
$\mathcal{A}_{96}$	972	817	9	82	507	1982	749	78	814	77
$\mathcal{A}_{97}$	977	818	93	83	53	1983	758	84	819	79
$\mathcal{A}_{98}$	978	822	94	9	72	1988	764	87	821	8
$\mathcal{A}_{99}$	979	823	97	92	77	1993	767	94	887	83
$\mathcal{A}_{100}$	982	824	98	93	86	1998	823	99	99	99

Table 3: Chosen clean command audio files.

Appendix B: Grid Search and EA Parameter Values

To find the the values of the four parameters $|\mathcal{P}|$, $|\mathcal{E}|$, ϵ_{start}, and ϵ_M of Speech-EA (referred to in the Experimental Settings Section of the article), we performed a grid search with the 256 parameter combination defined in Table 4:

$\lvert\mathcal{P}\rvert$	50	75	100	150
$\lvert\mathcal{E}\rvert/\lvert\mathcal{P}\rvert$	10 %	20 %	30 %	40 %
ϵ_{start}	0.9	0.7	0.5	0.25
ϵ_M	0.9	0.7	0.5	0.25

Table 4: Grid search combination settings

Concretely, the grid search consisted of running the EA-based attack for each combination of parameters from Table 4 against two particular audio samples (4.`wav` and 8.`wav`, from the command word `"LEFT"`) with a maximal number of 1,000 iterations. All runs used the same initial seed (42) for the random number generator.

Figure 6 provides a graphical illustration of the result of this grid search. The number i in Figure 6 represents the i^{th} combination of parameters satisfying two conditions: (1) Speech-EA terminates quickly and successfully on both audio samples, i.e. within at most 200 iterations for both audio samples; (2) the loss $\mathcal{L}_P$ remains moderate (below 15) for both audio samples. Then the average over the two audio samples considered determines the position of i in Figure 6. The colour of i indicates the range of the number of iterations required until success. The closer a point i is to the x-axis, the better (in terms of audio quality) this parameter set performs.

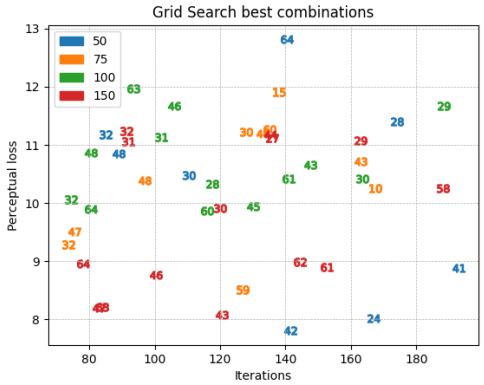

Fig. 6: Grid search for different parameter combinations as given in Table 4 for EA-based algorithm.

The 42$^{\text{nd}}$ combination (this 42 is not related to the random seed value!) achieves the lowest $\mathcal{L}_{\text{P}}$. The values of the corresponding parameters are $|\mathcal{P}| = 50$, $|\mathcal{E}| = 15$ (i.e., 30% of $|\mathcal{P}|$), $\epsilon_M = 0.7$ and $\epsilon_{start} = 0.5$. These are the values kept for the experiments.

	50		75		100		150	
	perceptual_loss	epochs	perceptual_loss	epochs	perceptual_loss	epochs	perceptual_loss	epochs
0.1_0.25_0.25	0.00	0	7.46	716	10.06	868	0.00	0
0.1_0.25_0.5	0.00	0	0.00	0	0.00	0	6.12	449
0.1_0.25_0.7	15.89	822	8.55	454	15.09	787	8.90	673
0.1_0.25_0.9	9.81	491	11.81	615	14.36	658	12.56	523
0.1_0.5_0.25	5.76	271	14.07	481	10.48	400	10.50	432
0.1_0.5_0.5	7.95	245	8.62	230	11.18	292	8.69	404
0.1_0.5_0.7	10.93	369	9.04	244	9.59	220	11.24	269
0.1_0.5_0.9	10.01	182	9.91	206	10.36	222	11.72	236
0.1_0.7_0.25	8.58	238	12.23	268	7.82	206	7.82	200
0.1_0.7_0.5	9.50	160	11.19	177	8.67	226	12.40	173
0.1_0.7_0.7	8.73	183	7.18	111	9.60	137	9.39	122
0.1_0.7_0.9	11.02	128	11.28	140	10.49	174	8.83	156
0.1_0.9_0.25	10.30	132	6.62	159	9.61	170	10.56	218
0.1_0.9_0.5	11.67	141	10.02	108	8.59	91	9.51	123
0.1_0.9_0.7	9.05	88	9.43	129	9.01	115	8.89	81
0.1_0.9_0.9	8.32	80	7.53	72	10.12	107	8.09	67
0.2_0.25_0.25	0.00	0	0.00	0	0.00	0	8.24	646
0.2_0.25_0.5	19.01	811	13.91	668	13.41	793	8.98	452
0.2_0.25_0.7	22.61	821	8.31	472	17.22	704	16.97	649
0.2_0.25_0.9	6.74	394	10.47	439	8.88	486	12.01	469
0.2_0.5_0.25	9.32	352	5.75	234	9.04	396	8.00	268
0.2_0.5_0.5	8.04	297	6.81	270	6.45	211	7.84	244
0.2_0.5_0.7	9.87	214	7.67	162	9.73	214	6.31	272
0.2_0.5_0.9	5.81	152	8.59	170	8.73	150	8.31	149
0.2_0.7_0.25	6.56	274	6.48	175	8.72	234	7.22	274
0.2_0.7_0.5	10.08	259	7.55	141	9.93	215	11.47	177
0.2_0.7_0.7	11.24	165	7.87	111	9.74	146	8.20	103
0.2_0.7_0.9	10.85	155	6.84	88	7.95	113	8.58	105
0.2_0.9_0.25	7.19	96	7.55	150	9.78	185	7.64	192
0.2_0.9_0.5	11.44	142	9.10	112	9.50	186	7.69	102
0.2_0.9_0.7	9.63	91	8.49	92	8.63	86	7.21	76
0.2_0.9_0.9	9.02	76	8.75	65	8.80	75	8.95	77
0.3_0.25_0.25	5.74	624	7.68	546	7.82	596	4.84	367
0.3_0.25_0.5	7.51	674	6.43	431	0.00	0	18.74	756
0.3_0.25_0.7	13.47	976	8.96	839	9.40	490	8.54	405
0.3_0.25_0.9	0.00	0	8.81	345	18.14	564	7.05	229
0.3_0.5_0.25	10.73	433	9.45	300	7.03	278	6.35	266
0.3_0.5_0.5	10.12	446	9.34	265	8.30	259	6.34	333
0.3_0.5_0.7	9.96	195	5.41	172	8.94	312	8.80	176
0.3_0.5_0.9	8.79	179	9.92	213	9.30	159	6.41	146
0.3_0.7_0.25	6.22	198	5.20	146	8.60	192	6.63	201
0.3_0.7_0.5	**6.48**	**150**	**11.11**	**161**	**7.08**	**180**	**7.93**	**191**
0.3_0.7_0.7	9.63	117	9.94	140	8.89	114	8.79	120
0.3_0.7_0.9	9.59	109	6.64	89	8.60	104	7.85	106
0.3_0.9_0.25	6.23	132	7.64	133	9.39	153	6.37	125
0.3_0.9_0.5	8.59	139	7.75	119	9.00	105	8.69	101
0.3_0.9_0.7	9.41	97	8.43	79	7.61	91	8.47	89
0.3_0.9_0.9	9.70	72	9.32	91	7.86	57	9.09	81
0.4_0.25_0.25	0.00	0	0.00	0	0.00	0	11.52	627
0.4_0.25_0.5	0.00	0	0.00	0	11.87	540	8.99	574
0.4_0.25_0.7	8.17	880	0.00	0	7.49	371	17.60	556
0.4_0.25_0.9	11.93	913	7.18	468	6.88	208	9.84	348
0.4_0.5_0.25	9.54	589	12.22	431	4.86	213	5.05	204
0.4_0.5_0.5	12.95	596	8.29	346	7.04	229	9.98	354
0.4_0.5_0.7	10.01	287	9.82	258	6.96	182	9.00	262
0.4_0.5_0.9	15.10	328	11.23	244	8.16	151	7.07	119
0.4_0.7_0.25	10.18	392	9.39	236	5.83	163	5.48	168
0.4_0.7_0.5	11.38	244	12.17	228	9.65	179	8.85	191
0.4_0.7_0.7	11.36	243	6.61	128	9.25	123	7.80	110
0.4_0.7_0.9	10.20	167	10.41	155	6.71	74	7.57	86
0.4_0.9_0.25	12.29	259	7.61	164	8.07	146	7.00	151
0.4_0.9_0.5	12.16	178	9.67	182	8.04	97	6.75	197
0.4_0.9_0.7	10.08	210	11.56	140	9.11	94	6.38	62
0.4_0.9_0.9	11.37	147	10.64	145	7.97	70	8.19	67

Fig. 7: Grid Search combinations for 4.`wav`; Combination 42 is highlighted (in red).

	50		75		100		150	
	perceptual_loss	epochs	perceptual_loss	epochs	perceptual_loss	epochs	perceptual_loss	epochs
0.1_0.25_0.25	0.00	0	0.00	0	0.00	0	0.00	0
0.1_0.25_0.5	0.00	0	0.00	0	0.00	0	0.00	0
0.1_0.25_0.7	0.00	0	0.00	0	0.00	0	0.00	0
0.1_0.25_0.9	0.00	0	0.00	0	0.00	0	13.05	628
0.1_0.5_0.25	0.00	0	11.11	337	0.00	0	19.93	520
0.1_0.5_0.5	26.57	484	13.86	532	26.98	837	11.79	303
0.1_0.5_0.7	23.45	983	0.00	0	41.70	678	21.53	427
0.1_0.5_0.9	13.88	293	15.50	550	29.22	448	22.60	461
0.1_0.7_0.25	15.39	310	23.72	358	20.63	381	12.49	587
0.1_0.7_0.5	21.43	280	9.34	158	12.44	221	34.62	365
0.1_0.7_0.7	19.99	187	13.58	259	18.95	282	16.35	168
0.1_0.7_0.9	38.41	360	15.73	160	17.39	150	18.57	172
0.1_0.9_0.25	27.88	285	15.54	163	17.55	337	21.02	222
0.1_0.9_0.5	17.72	194	18.57	158	21.30	158	24.77	207
0.1_0.9_0.7	15.27	108	14.40	147	15.41	97	25.28	190
0.1_0.9_0.9	21.46	205	17.90	111	29.73	150	18.25	111
0.2_0.25_0.25	0.00	0	0.00	0	0.00	0	0.00	0
0.2_0.25_0.5	0.00	0	0.00	0	0.00	0	16.78	836
0.2_0.25_0.7	0.00	0	14.50	723	0.00	0	0.00	0
0.2_0.25_0.9	0.00	0	0.00	0	0.00	0	0.00	0
0.2_0.5_0.25	11.45	455	11.07	420	30.71	866	13.51	363
0.2_0.5_0.5	21.90	747	13.01	311	0.00	0	12.75	603
0.2_0.5_0.7	20.60	576	13.81	260	10.69	191	12.56	322
0.2_0.5_0.9	10.22	182	38.95	474	15.60	574	13.14	261
0.2_0.7_0.25	23.11	315	18.69	587	18.09	249	17.11	386
0.2_0.7_0.5	20.98	413	16.18	271	14.60	178	12.30	319
0.2_0.7_0.7	17.59	297	18.08	206	15.49	267	14.05	169
0.2_0.7_0.9	11.95	193	15.33	181	12.70	122	13.43	250
0.2_0.9_0.25	16.11	151	12.18	204	13.57	192	14.50	134
0.2_0.9_0.5	9.52	79	13.34	144	11.33	141	12.13	138
0.2_0.9_0.7	27.36	210	16.60	167	13.65	118	14.93	108
0.2_0.9_0.9	13.34	94	9.84	82	11.32	74	13.53	106
0.3_0.25_0.25	0.00	0	0.00	0	0.00	0	0.00	0
0.3_0.25_0.5	0.00	0	0.00	0	0.00	0	0.00	0
0.3_0.25_0.7	0.00	0	0.00	0	0.00	0	0.00	0
0.3_0.25_0.9	0.00	0	0.00	0	0.00	0	0.00	0
0.3_0.5_0.25	0.00	0	0.00	0	16.66	865	10.16	761
0.3_0.5_0.5	12.15	579	15.30	426	20.79	342	10.41	276
0.3_0.5_0.7	24.02	454	17.73	316	21.68	455	10.88	367
0.3_0.5_0.9	9.35	223	10.64	269	11.70	259	10.22	249
0.3_0.7_0.25	11.56	188	22.23	423	17.66	331	20.67	350
0.3_0.7_0.5	**9.14**	**133**	**11.09**	**205**	**11.43**	**224**	**13.96**	**289**
0.3_0.7_0.7	15.05	180	11.49	186	12.42	181	7.36	121
0.3_0.7_0.9	18.08	201	13.69	258	16.70	203	14.51	165
0.3_0.9_0.25	17.82	181	18.16	198	10.50	107	18.82	150
0.3_0.9_0.5	15.46	126	14.63	147	14.34	107	8.83	100
0.3_0.9_0.7	15.75	102	10.60	72	19.12	120	7.92	77
0.3_0.9_0.9	12.01	106	11.47	103	13.86	104	15.56	129
0.4_0.25_0.25	0.00	0	0.00	0	13.78	853	0.00	0
0.4_0.25_0.5	0.00	0	0.00	0	14.63	823	0.00	0
0.4_0.25_0.7	0.00	0	0.00	0	10.02	694	0.00	0
0.4_0.25_0.9	0.00	0	0.00	0	0.00	0	0.00	0
0.4_0.5_0.25	0.00	0	44.87	992	18.62	454	20.91	914
0.4_0.5_0.5	31.11	707	18.93	508	11.62	315	11.19	546
0.4_0.5_0.7	14.21	876	26.76	468	14.35	352	11.54	309
0.4_0.5_0.9	27.94	424	24.05	514	16.95	288	12.27	257
0.4_0.7_0.25	17.35	562	16.65	390	11.56	406	11.78	229
0.4_0.7_0.5	16.22	356	19.47	266	22.42	221	11.65	185
0.4_0.7_0.7	19.49	303	10.42	126	12.58	200	12.03	203
0.4_0.7_0.9	27.25	230	12.15	115	13.01	158	15.15	162
0.4_0.9_0.25	15.91	215	25.52	257	12.77	136	10.79	154
0.4_0.9_0.5	22.36	184	32.62	191	16.00	113	11.23	92
0.4_0.9_0.7	23.32	175	15.80	137	14.83	93	10.06	106
0.4_0.9_0.9	14.25	134	26.18	160	11.83	91	9.72	89

Fig. 8: Grid Search combinations for 8.`wav`; Combination 42 is highlighted (in red).

Appendix C: Mel Spectrograms

Figures 9, 10, 11,12, 13, 14, 15, 16, 17 and 18 show representative examples of the waveform (top) and the Mel-spectrogram (bottom) representation of an adversarial audio and a clean audio for each command word.

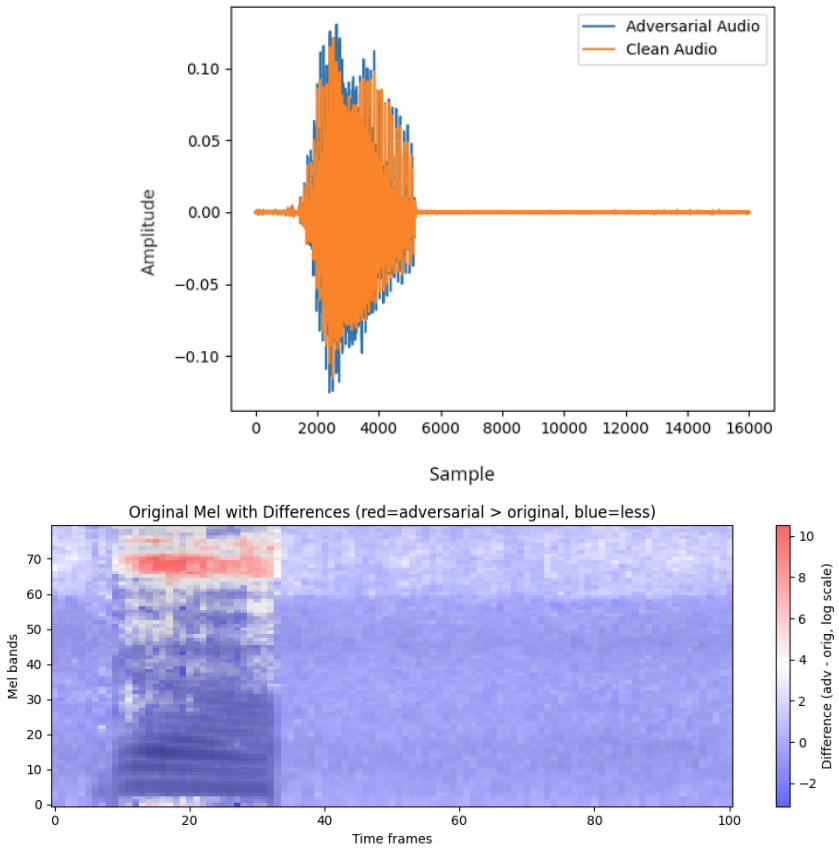

Fig. 9: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"YES"` (file 489.`wav`).

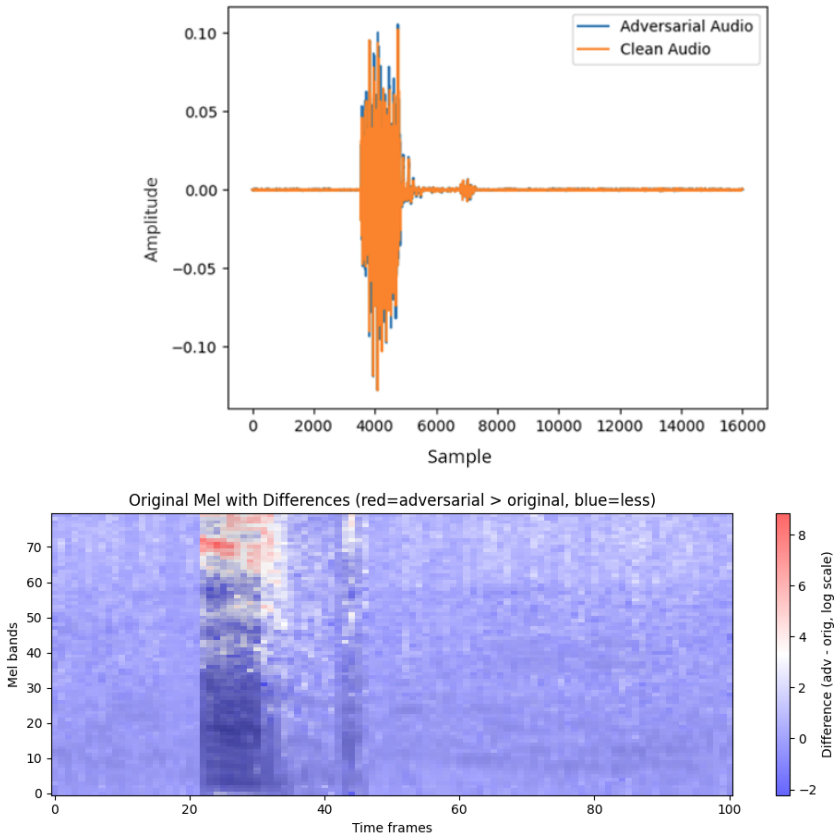

Fig. 10: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"UP"` (file 479.`wav`).

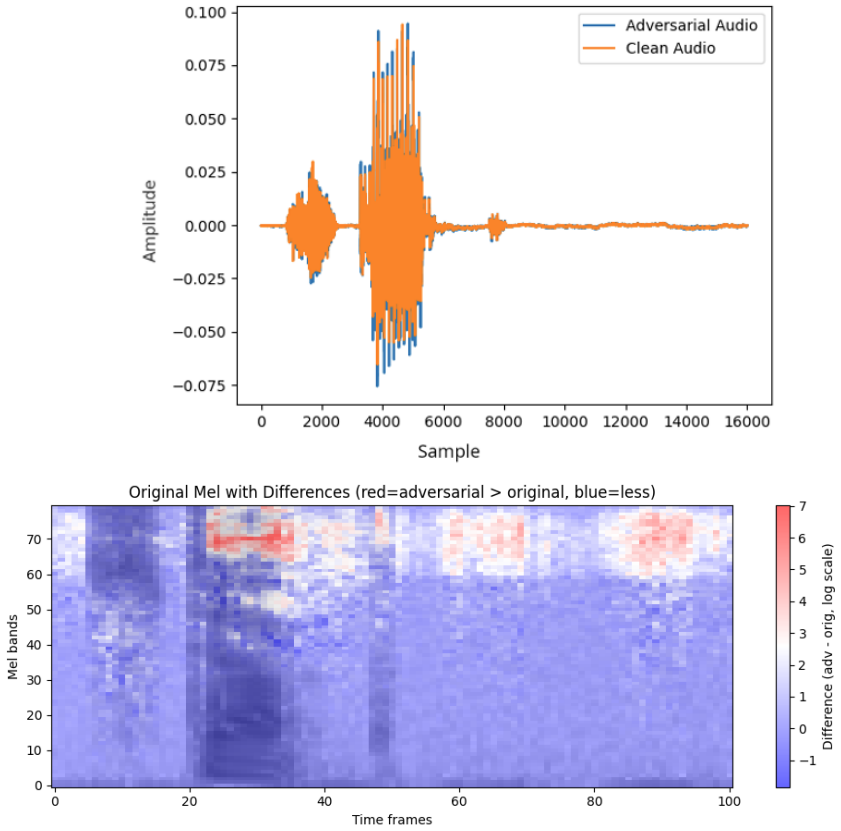

Fig. 11: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"STOP"` (file 473.`wav`).

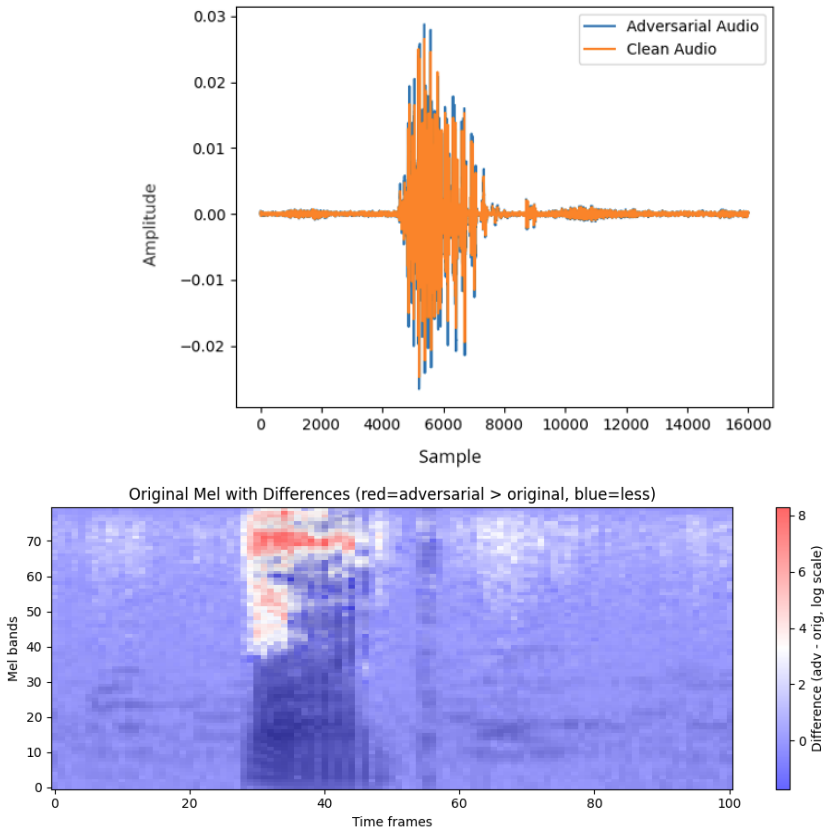

Fig. 12: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"RIGHT"` (file 449.`wav`).

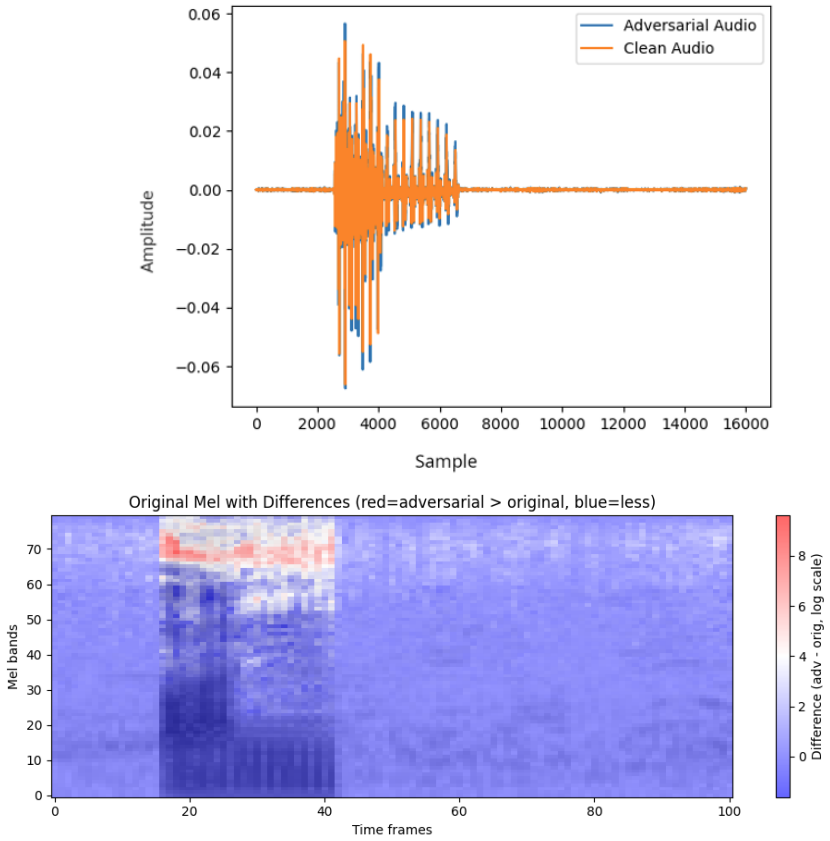

Fig. 13: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"ON"` (file 1944.`wav`).

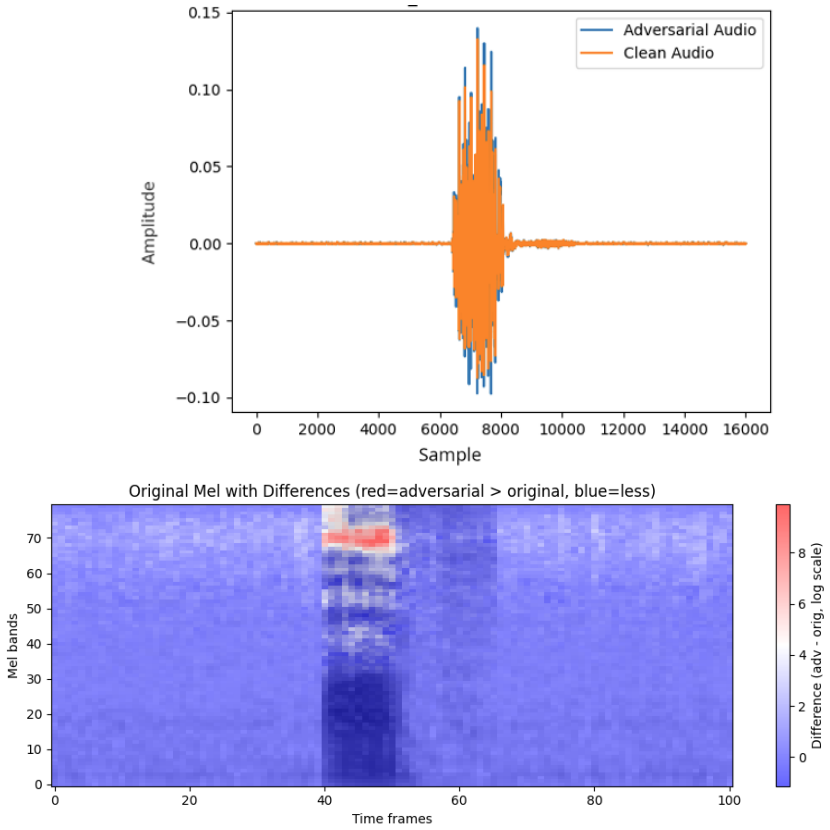

Fig. 14: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"OFF"` (file 464.`wav`).

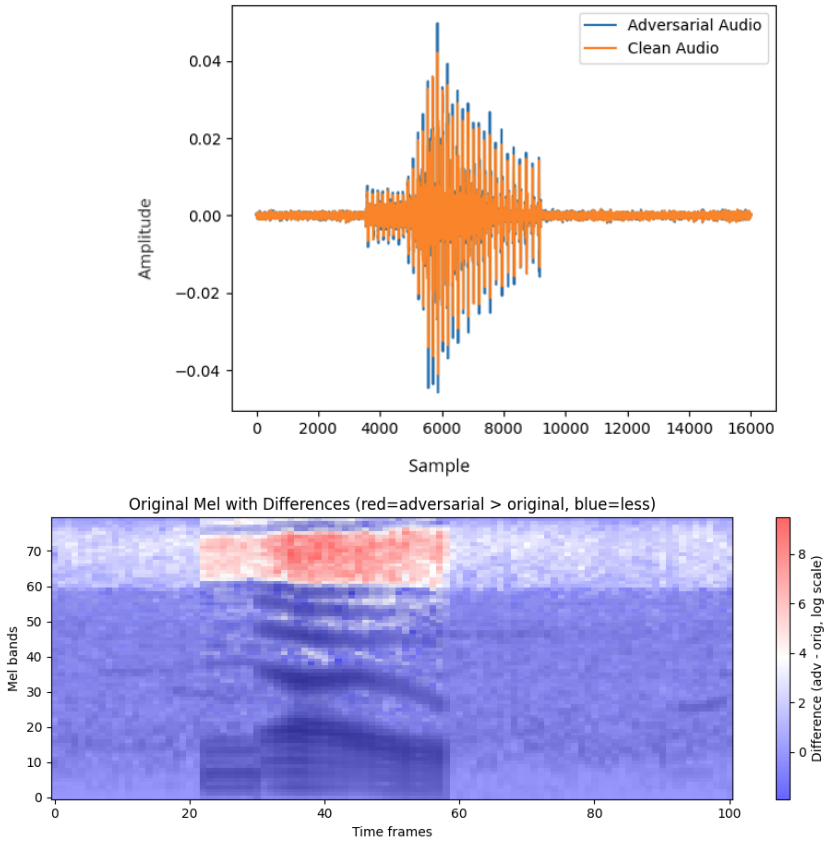

Fig. 15: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"NO"` (file 29.`wav`).

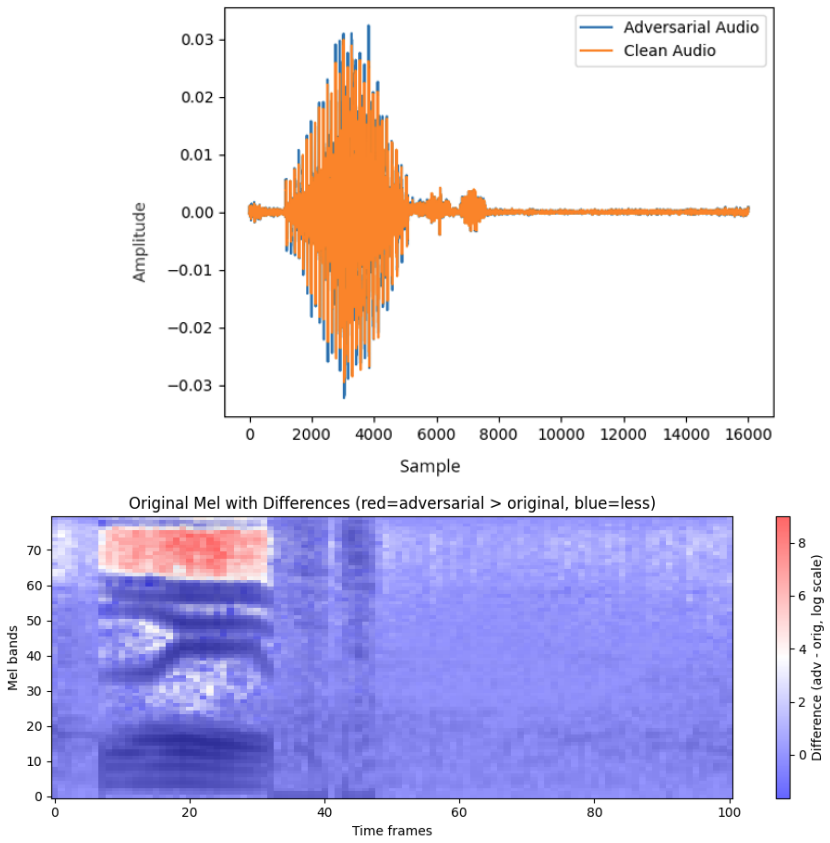

Fig. 16: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"LEFT"` (file 369.wav).

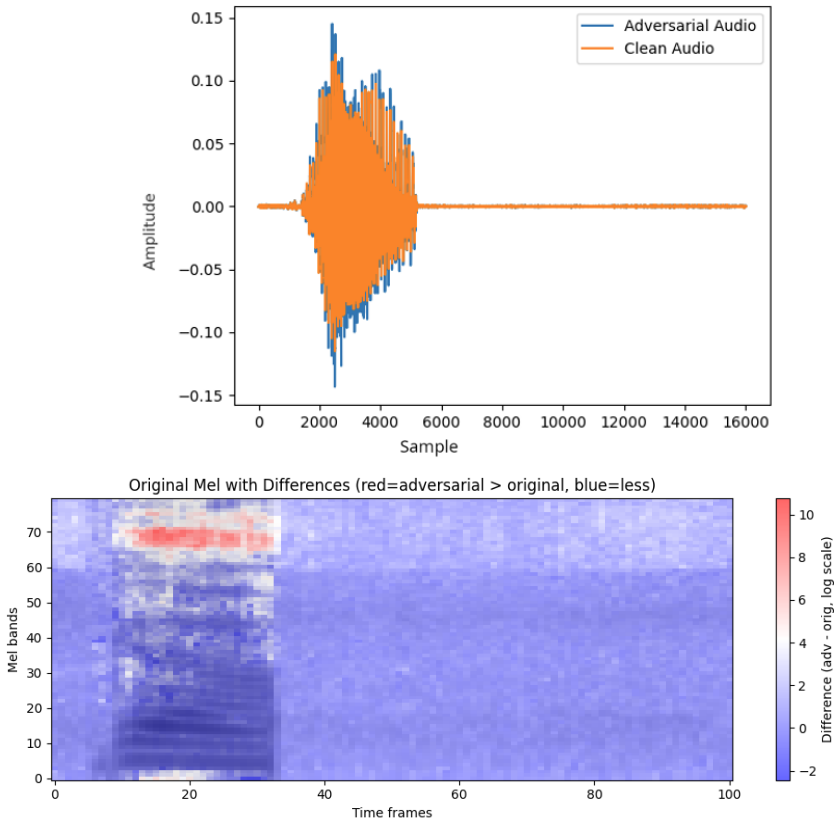

Fig. 17: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"GO"` (file 489.wav).

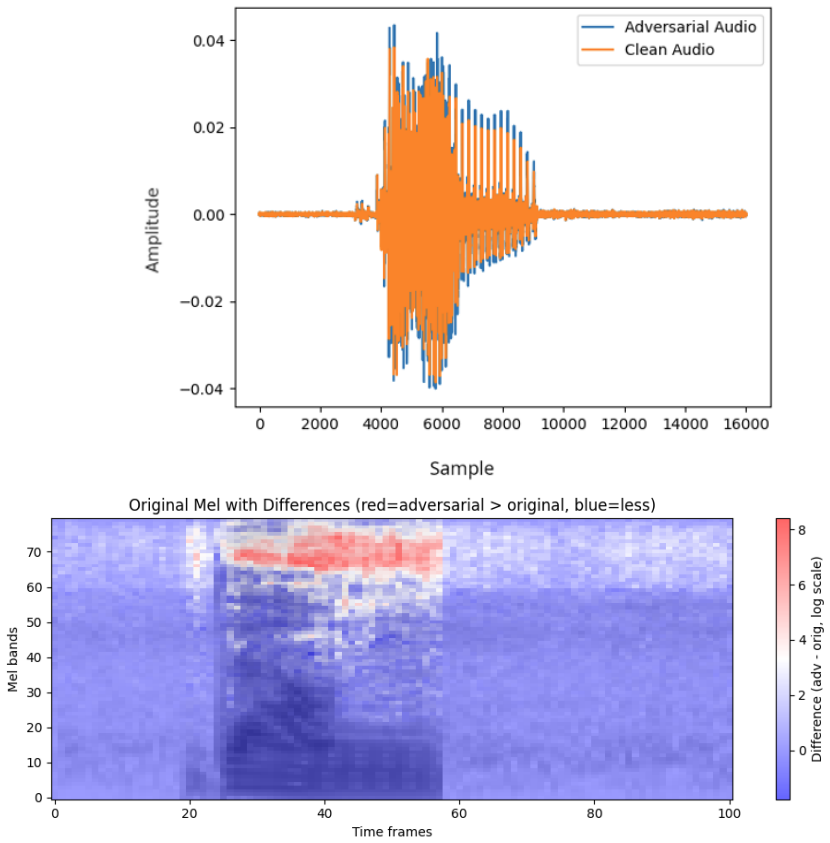

Fig. 18: The waveform of x and x_{adv} (top graph) and Mel-spectrogram representation showing the noise added to x (bottom graph, where red means larger difference) for the command `"DOWN"` (file 969.`wav`).