

Comparison of Discrete Transforms For DNN-Based Speech Enhancement

by

Wissam A. Jassim and Naomi Harte

wissam.a.jassim@gmail.com, nharte@tcd.ie

SigmaMedia, ADAPT Centre, School of Engineering, Trinity College Dublin, Ireland

Jul-21

Quality and intelligibility scores for DNN-based speech enhancement using different discrete transforms: DFT, DCT, DST, DTT, DKT, DTKT, DHT, DHDT, DTCS, and DCST. The tested speech signals were distorted by four noise types (babble, car, restaurant, and street) at four SNR levels (0, 5, 10, and 15 dB). The best results are highlighted with a color shading.

Table 2. List of objective measures with their abbreviations.

Abbr.	Objective Measure
PESQ	Quality: Perceptual evaluation of speech quality [11]
OVL	Quality: Overall speech quality [9]
CSII	Ineligibility: Coherence speech intelligibility index [12]
NCM	Ineligibility: Ineligibility in noisy conditions [13]

Measure:		PESQ				OVL				CSII				NCM			
Noise Level:		0 dB	5 dB	10 dB	15 dB	0 dB	5 dB	10 dB	15 dB	0 dB	5 dB	10 dB	15 dB	0 dB	5 dB	10 dB	15 dB
Noise Type	Method																
Babble	Noisy	1.74	2.04	2.34	2.66	2.07	2.50	2.92	3.32	0.49	0.82	0.96	0.99	0.54	0.73	0.89	0.97
	DFT	1.79	2.05	2.24	2.36	1.94	2.31	2.57	2.74	0.44	0.76	0.92	0.96	0.59	0.70	0.77	0.79
	DCT	1.75	2.02	2.19	2.31	1.90	2.25	2.49	2.66	0.39	0.70	0.89	0.95	0.58	0.69	0.76	0.78
	DST	1.75	2.01	2.19	2.31	1.90	2.25	2.50	2.67	0.39	0.71	0.89	0.95	0.58	0.69	0.76	0.78
	DTT	1.77	2.05	2.26	2.41	1.92	2.29	2.57	2.77	0.40	0.73	0.90	0.96	0.60	0.72	0.79	0.81
	DKT	1.75	2.00	2.21	2.36	1.86	2.21	2.49	2.67	0.39	0.71	0.89	0.95	0.59	0.71	0.77	0.80
	DTKT	1.64	1.91	2.13	2.29	1.71	2.06	2.36	2.55	0.34	0.66	0.86	0.93	0.58	0.71	0.78	0.81
	DHT	1.62	1.86	2.04	2.16	1.67	1.99	2.24	2.39	0.34	0.64	0.83	0.90	0.57	0.69	0.75	0.77
	DHDT	1.71	1.96	2.16	2.30	1.78	2.13	2.41	2.59	0.36	0.68	0.87	0.94	0.58	0.71	0.78	0.81
	DTCS	1.77	2.05	2.27	2.42	1.92	2.29	2.58	2.77	0.41	0.73	0.90	0.96	0.60	0.72	0.79	0.82
	DCST	1.77	2.05	2.26	2.41	1.92	2.29	2.57	2.77	0.41	0.73	0.90	0.96	0.60	0.72	0.79	0.81
Car	Noisy	1.66	1.91	2.21	2.54	1.97	2.37	2.78	3.20	0.45	0.78	0.95	0.99	0.61	0.79	0.90	0.97
	DFT	1.83	2.05	2.22	2.35	2.03	2.33	2.56	2.73	0.52	0.78	0.92	0.96	0.67	0.73	0.77	0.79
	DCT	1.78	1.99	2.17	2.30	1.97	2.25	2.49	2.66	0.45	0.72	0.89	0.95	0.65	0.71	0.76	0.78
	DST	1.77	1.99	2.17	2.30	1.97	2.25	2.49	2.67	0.45	0.72	0.89	0.95	0.65	0.71	0.76	0.78
	DTT	1.76	2.02	2.22	2.39	1.95	2.27	2.54	2.75	0.46	0.73	0.90	0.95	0.66	0.74	0.79	0.81
	DKT	1.70	1.94	2.15	2.32	1.84	2.16	2.43	2.64	0.41	0.68	0.88	0.94	0.64	0.72	0.77	0.80
	DTKT	1.59	1.83	2.06	2.26	1.63	1.97	2.27	2.51	0.36	0.64	0.85	0.93	0.64	0.73	0.78	0.81
	DHT	1.58	1.80	2.00	2.14	1.62	1.93	2.19	2.37	0.36	0.62	0.82	0.90	0.63	0.70	0.75	0.77
	DHDT	1.66	1.89	2.10	2.27	1.74	2.07	2.35	2.56	0.38	0.67	0.86	0.94	0.63	0.73	0.78	0.81
	DTCS	1.76	2.02	2.22	2.39	1.94	2.27	2.54	2.75	0.46	0.73	0.90	0.96	0.66	0.74	0.79	0.82
	DCST	1.76	2.02	2.22	2.39	1.94	2.27	2.53	2.74	0.46	0.73	0.90	0.95	0.66	0.73	0.79	0.81
Restaurant	Noisy	1.79	2.02	2.39	2.68	2.16	2.52	2.98	3.34	0.56	0.83	0.97	0.99	0.53	0.72	0.86	0.96
	DFT	1.82	2.04	2.25	2.37	1.99	2.29	2.57	2.74	0.48	0.77	0.92	0.96	0.57	0.69	0.76	0.79
	DCT	1.79	1.99	2.21	2.31	1.95	2.23	2.51	2.66	0.42	0.71	0.89	0.95	0.56	0.68	0.75	0.78
	DST	1.79	1.99	2.20	2.31	1.96	2.23	2.51	2.67	0.42	0.72	0.89	0.95	0.56	0.68	0.75	0.78
	DTT	1.82	2.04	2.29	2.42	1.98	2.28	2.59	2.76	0.44	0.74	0.91	0.96	0.57	0.70	0.78	0.81
	DKT	1.77	1.98	2.22	2.36	1.90	2.20	2.50	2.68	0.42	0.71	0.89	0.95	0.56	0.69	0.76	0.80
	DTKT	1.70	1.90	2.15	2.30	1.77	2.06	2.37	2.56	0.38	0.66	0.87	0.94	0.54	0.69	0.77	0.81
	DHT	1.66	1.84	2.07	2.17	1.71	1.97	2.25	2.39	0.35	0.62	0.83	0.90	0.54	0.68	0.74	0.77
	DHDT	1.74	1.93	2.18	2.30	1.83	2.11	2.42	2.59	0.38	0.67	0.88	0.94	0.56	0.69	0.77	0.81
	DTCS	1.82	2.03	2.29	2.42	1.98	2.28	2.59	2.77	0.45	0.74	0.91	0.96	0.57	0.70	0.78	0.82
	DCST	1.82	2.04	2.28	2.42	1.98	2.27	2.59	2.77	0.44	0.74	0.91	0.96	0.57	0.70	0.78	0.81
Street	Noisy	1.58	1.91	2.27	2.55	1.93	2.39	2.84	3.19	0.49	0.77	0.96	0.99	0.53	0.72	0.87	0.95
	DFT	1.78	2.02	2.22	2.33	1.97	2.29	2.54	2.71	0.49	0.76	0.92	0.96	0.60	0.72	0.77	0.79
	DCT	1.73	1.97	2.18	2.28	1.93	2.24	2.49	2.64	0.44	0.71	0.89	0.94	0.58	0.71	0.76	0.78
	DST	1.74	1.98	2.17	2.28	1.93	2.25	2.49	2.65	0.44	0.72	0.89	0.95	0.58	0.71	0.76	0.78
	DTT	1.76	2.01	2.24	2.38	1.94	2.27	2.56	2.73	0.45	0.73	0.91	0.95	0.59	0.73	0.79	0.81
	DKT	1.71	1.96	2.20	2.32	1.86	2.20	2.48	2.65	0.41	0.69	0.89	0.95	0.57	0.71	0.77	0.80
	DTKT	1.60	1.85	2.09	2.26	1.68	2.03	2.31	2.51	0.37	0.66	0.88	0.94	0.56	0.72	0.78	0.81
	DHT	1.59	1.81	2.02	2.14	1.65	1.96	2.21	2.36	0.36	0.62	0.84	0.90	0.56	0.70	0.75	0.77
	DHDT	1.64	1.89	2.13	2.27	1.76	2.09	2.38	2.56	0.40	0.68	0.88	0.94	0.57	0.72	0.78	0.81
	DTCS	1.75	2.01	2.25	2.38	1.93	2.27	2.56	2.74	0.44	0.73	0.90	0.95	0.59	0.73	0.79	0.81
	DCST	1.75	2.00	2.25	2.38	1.93	2.26	2.55	2.73	0.44	0.73	0.90	0.95	0.59	0.73	0.78	0.81
Mean	Noisy	1.69	1.97	2.30	2.61	2.03	2.44	2.88	3.26	0.50	0.80	0.96	0.99	0.55	0.74	0.88	0.96
	DFT	1.81	2.04	2.23	2.35	1.98	2.30	2.56	2.73	0.48	0.77	0.92	0.96	0.61	0.71	0.77	0.79
	DCT	1.76	1.99	2.19	2.30	1.94	2.24	2.50	2.66	0.43	0.71	0.89	0.95	0.59	0.70	0.75	0.78
	DST	1.76	1.99	2.18	2.30	1.94	2.24	2.50	2.66	0.43	0.71	0.89	0.95	0.59	0.70	0.75	0.78
	DTT	1.78	2.03	2.25	2.40	1.95	2.28	2.56	2.75	0.44	0.73	0.90	0.96	0.60	0.72	0.78	0.81
	DKT	1.73	1.97	2.20	2.34	1.86	2.19	2.47	2.66	0.41	0.70	0.89	0.95	0.59	0.71	0.77	0.80
	DTKT	1.63	1.87	2.11	2.28	1.70	2.03	2.33	2.53	0.36	0.65	0.87	0.94	0.58	0.71	0.78	0.81
	DHT	1.61	1.83	2.03	2.15	1.66	1.96	2.22	2.38	0.35	0.63	0.83	0.90	0.57	0.69	0.75	0.77
	DHDT	1.69	1.92	2.14	2.29	1.78	2.10	2.39	2.58	0.38	0.67	0.87	0.94	0.58	0.71	0.78	0.81
	DTCS	1.77	2.03	2.26	2.40	1.94	2.28	2.57	2.76	0.44	0.73	0.90	0.96	0.60	0.72	0.79	0.82
	DCST	1.78	2.03	2.25	2.40	1.94	2.27	2.56	2.75	0.44	0.73	0.90	0.95	0.60	0.72	0.78	0.81

Table 1:
results for the FNN-
based model

Measure:		PESQ				OVL				CSII				NCM			
Noise Level:		0 dB	5 dB	10 dB	15 dB	0 dB	5 dB	10 dB	15 dB	0 dB	5 dB	10 dB	15 dB	0 dB	5 dB	10 dB	15 dB
Noise Type	Method																
Babble	Noisy	1.74	2.04	2.34	2.66	2.07	2.50	2.92	3.32	0.49	0.82	0.96	0.99	0.54	0.73	0.89	0.97
	DFT	1.92	2.24	2.46	2.58	2.16	2.60	2.89	3.02	0.51	0.83	0.95	0.98	0.61	0.77	0.85	0.87
	DCT	1.93	2.28	2.54	2.68	2.08	2.52	2.84	2.99	0.51	0.83	0.95	0.98	0.59	0.76	0.86	0.88
	DST	1.94	2.28	2.56	2.69	2.11	2.58	2.92	3.08	0.50	0.83	0.95	0.98	0.60	0.76	0.86	0.88
	DTT	1.90	2.23	2.51	2.69	2.07	2.51	2.84	3.05	0.49	0.82	0.95	0.98	0.61	0.77	0.87	0.90
	DKT	1.85	2.14	2.38	2.52	2.02	2.42	2.70	2.85	0.48	0.80	0.93	0.97	0.60	0.76	0.84	0.87
	DTKT	1.77	2.04	2.28	2.47	1.87	2.23	2.54	2.76	0.40	0.73	0.89	0.95	0.60	0.74	0.82	0.85
	DHT	1.71	1.95	2.13	2.24	1.78	2.08	2.31	2.45	0.39	0.68	0.84	0.90	0.60	0.74	0.81	0.82
	DHDT	1.76	2.01	2.20	2.33	1.85	2.18	2.42	2.57	0.39	0.72	0.88	0.94	0.60	0.74	0.82	0.84
	DTCS	1.90	2.24	2.53	2.68	2.09	2.53	2.87	3.06	0.49	0.83	0.95	0.98	0.61	0.77	0.87	0.90
	DCST	1.91	2.23	2.52	2.69	2.08	2.52	2.87	3.08	0.49	0.82	0.95	0.98	0.61	0.77	0.86	0.90
Car	Noisy	1.66	1.91	2.21	2.54	1.97	2.37	2.78	3.20	0.45	0.78	0.95	0.99	0.61	0.79	0.90	0.97
	DFT	1.93	2.22	2.45	2.58	2.22	2.60	2.87	3.02	0.56	0.84	0.95	0.98	0.70	0.81	0.85	0.87
	DCT	1.95	2.28	2.53	2.69	2.16	2.57	2.85	3.01	0.57	0.85	0.95	0.98	0.69	0.81	0.86	0.88
	DST	1.97	2.30	2.55	2.69	2.22	2.64	2.94	3.09	0.58	0.85	0.95	0.98	0.69	0.81	0.86	0.89
	DTT	1.90	2.21	2.47	2.67	2.11	2.51	2.79	3.03	0.54	0.84	0.95	0.98	0.70	0.82	0.87	0.90
	DKT	1.84	2.09	2.33	2.49	2.05	2.39	2.67	2.84	0.50	0.80	0.93	0.97	0.68	0.80	0.85	0.87
	DTKT	1.72	1.97	2.20	2.42	1.83	2.17	2.47	2.71	0.43	0.72	0.89	0.95	0.68	0.78	0.82	0.85
	DHT	1.66	1.88	2.07	2.20	1.75	2.03	2.25	2.41	0.42	0.69	0.83	0.90	0.67	0.77	0.81	0.83
	DHDT	1.71	1.95	2.14	2.30	1.81	2.12	2.35	2.53	0.43	0.71	0.87	0.94	0.67	0.77	0.82	0.84
	DTCS	1.90	2.22	2.50	2.68	2.13	2.53	2.86	3.06	0.55	0.84	0.95	0.98	0.70	0.82	0.87	0.90
	DCST	1.90	2.21	2.49	2.68	2.12	2.54	2.85	3.06	0.54	0.84	0.95	0.98	0.70	0.82	0.87	0.90
Restaurant	Noisy	1.79	2.02	2.39	2.68	2.16	2.52	2.98	3.34	0.56	0.83	0.97	0.99	0.53	0.72	0.86	0.96
	DFT	1.93	2.18	2.45	2.58	2.19	2.55	2.86	3.03	0.54	0.83	0.95	0.98	0.59	0.75	0.83	0.87
	DCT	1.95	2.22	2.51	2.67	2.11	2.46	2.80	2.99	0.54	0.82	0.95	0.98	0.58	0.74	0.83	0.88
	DST	1.96	2.22	2.52	2.68	2.15	2.53	2.88	3.06	0.54	0.82	0.95	0.98	0.58	0.74	0.83	0.88
	DTT	1.93	2.18	2.49	2.70	2.12	2.46	2.81	3.04	0.53	0.81	0.95	0.98	0.58	0.75	0.84	0.89
	DKT	1.86	2.11	2.36	2.52	2.04	2.39	2.68	2.85	0.51	0.79	0.93	0.97	0.57	0.73	0.83	0.87
	DTKT	1.81	2.01	2.29	2.47	1.92	2.22	2.54	2.76	0.42	0.71	0.89	0.95	0.55	0.72	0.81	0.85
	DHT	1.73	1.92	2.14	2.24	1.81	2.06	2.31	2.44	0.41	0.66	0.84	0.90	0.56	0.71	0.79	0.82
	DHDT	1.79	1.98	2.21	2.33	1.88	2.15	2.41	2.57	0.42	0.70	0.88	0.94	0.55	0.71	0.80	0.84
	DTCS	1.93	2.18	2.49	2.69	2.14	2.48	2.83	3.05	0.53	0.82	0.95	0.98	0.58	0.75	0.84	0.89
	DCST	1.94	2.18	2.50	2.70	2.13	2.48	2.84	3.07	0.53	0.82	0.95	0.98	0.58	0.75	0.84	0.89
Street	Noisy	1.58	1.91	2.27	2.55	1.93	2.39	2.84	3.19	0.49	0.77	0.96	0.99	0.53	0.72	0.87	0.95
	DFT	1.88	2.18	2.43	2.54	2.15	2.56	2.84	2.98	0.57	0.82	0.95	0.98	0.64	0.78	0.85	0.87
	DCT	1.91	2.24	2.54	2.65	2.11	2.53	2.85	2.98	0.58	0.82	0.95	0.98	0.63	0.78	0.86	0.88
	DST	1.93	2.25	2.54	2.64	2.15	2.58	2.89	3.03	0.58	0.83	0.95	0.97	0.63	0.78	0.86	0.88
	DTT	1.87	2.20	2.49	2.62	2.06	2.47	2.80	2.97	0.54	0.81	0.95	0.98	0.62	0.78	0.87	0.89
	DKT	1.82	2.10	2.35	2.47	2.01	2.40	2.67	2.81	0.51	0.77	0.93	0.96	0.60	0.76	0.84	0.86
	DTKT	1.70	1.97	2.25	2.41	1.83	2.20	2.51	2.71	0.41	0.70	0.90	0.95	0.58	0.75	0.82	0.85
	DHT	1.65	1.89	2.09	2.20	1.75	2.05	2.28	2.41	0.41	0.66	0.84	0.90	0.57	0.74	0.80	0.82
	DHDT	1.68	1.94	2.17	2.29	1.78	2.12	2.38	2.53	0.39	0.69	0.88	0.94	0.56	0.74	0.81	0.84
	DTCS	1.87	2.20	2.50	2.64	2.07	2.50	2.83	3.00	0.55	0.81	0.95	0.98	0.62	0.78	0.87	0.89
	DCST	1.87	2.20	2.49	2.64	2.07	2.50	2.83	3.02	0.55	0.81	0.95	0.98	0.62	0.78	0.86	0.89
Mean	Noisy	1.69	1.97	2.30	2.61	2.03	2.44	2.88	3.26	0.50	0.80	0.96	0.99	0.55	0.74	0.88	0.96
	DFT	1.92	2.21	2.45	2.57	2.18	2.58	2.86	3.01	0.55	0.83	0.95	0.98	0.63	0.78	0.84	0.87
	DCT	1.93	2.26	2.53	2.67	2.11	2.52	2.84	2.99	0.55	0.83	0.95	0.98	0.62	0.77	0.85	0.88
	DST	1.95	2.26	2.54	2.68	2.16	2.58	2.91	3.07	0.55	0.83	0.95	0.98	0.62	0.78	0.85	0.88
	DTT	1.90	2.21	2.49	2.67	2.09	2.49	2.81	3.02	0.53	0.82	0.95	0.98	0.63	0.78	0.86	0.90
	DKT	1.84	2.11	2.35	2.50	2.03	2.40	2.68	2.84	0.50	0.79	0.93	0.97	0.61	0.76	0.84	0.87
	DTKT	1.75	2.00	2.26	2.44	1.86	2.20	2.51	2.73	0.42	0.72	0.89	0.95	0.60	0.75	0.82	0.85
	DHT	1.69	1.91	2.11	2.22	1.77	2.06	2.29	2.43	0.41	0.67	0.84	0.90	0.60	0.74	0.80	0.82
	DHDT	1.73	1.97	2.18	2.31	1.83	2.14	2.39	2.55	0.41	0.70	0.88	0.94	0.59	0.74	0.81	0.84
	DTCS	1.90	2.21	2.51	2.67	2.11	2.51	2.85	3.04	0.53	0.83	0.95	0.98	0.63	0.78	0.86	0.89
	DCST	1.90	2.21	2.50	2.68	2.10	2.51	2.85	3.06	0.53	0.82	0.95	0.98	0.63	0.78	0.86	0.89

Table 2:
results for the CNN-
based model