Table 1: Ablation Study on HQ-VoxCeleb dataset

Setting	**Similarity**		**Retrieval**				**Quality**
Ablation	cosine	L1	R@1	R@2	R@5	R@10	VFS
w/o L_1	0.310	21.74	3.47	7.26	18.36	31.56	16.84
w/o L_G	0.298	20.97	3.23	8.70	16.73	32.36	18.51
w/o L_C	0.310	18.13	5.24	8.45	17.30	34.28	19.16
w/o L_P	0.236	18.22	2.73	5.23	15.39	26.37	18.44
Baseline Encoder	0.309	18.47	5.54	9.26	19.18	34.75	18.89
Baseline Decoder	0.311	18.52	4.32	8.56	19.47	34.25	18.46
LQ. Dataset	0.267	19.94	1.96	4.21	10.67	21.34	16.73
Full Model	0.317	17.91	5.75	9.32	20.36	36.65	19.49