

S2S Models	Human				GPT-4o Realtime				Gemini-2.5-pro				Qwen-Omni-Turbo			
	Sem.	Para.	Ambi.	Ovrl.↑	Sem.	Para.	Ambi.	Ovrl.	Sem.	Para.	Ambi.	Ovrl.	Sem.	Para.	Ambi.	Ovrl.
Arena-style Evaluation																
Closed-source Models																
Doubao	1023	1038	1049	1037	1008	1029	1020	1019	1012	1027	1005	1015	985	1017	997	1000
Qwen-Omni-Turbo	1007	1044	1051	1034	994	1020	986	1000	1002	982	983	989	1012	1007	1009	1009
GPT-4o Realtime	1041	1029	1020	1030	1036	1005	1032	1024	1023	1022	1028	1025	1011	984	1017	1004
Open-source Models																
Step-Audio-Chat	1058	1054	1034	1049	1027	1011	1009	1016	1025	994	1013	1010	1007	1041	1050	1033
GLM-4- Voice	1010	1011	982	1001	1009	1004	1015	1010	1017	1023	1009	1016	1018	1001	1010	1010
VITA-Audio-Plus- Vanilla	1004	973	1001	993	1025	1004	1029	1019	995	999	1004	999	1035	994	1013	1014
MiniCPM-o 2.6	989	980	979	983	973	987	986	982	972	984	997	984	992	1008	994	998
Kimi-Audio	970	978	972	973	984	1048	997	1010	996	1032	1025	1018	990	999	976	988
Moshi	950	970	947	956	987	965	991	981	971	969	969	970	985	984	984	984
AnyGPT	942	941	931	938	956	924	960	947	977	980	990	982	1001	984	962	982
Human	1005	981	1034	1006	981	970	968	973	998	983	970	983	945	972	960	959
Rubric-based Evaluation																
Closed-source Models																
GPT-4o Realtime	88.59	73.75	69.73	77.38	88.06	76.01	67.80	77.33	73.76	64.34	70.31	69.50	84.85	80.33	86.51	84.01
Doubao	82.54	77.06	60.42	73.69	81.97	73.61	60.50	72.06	70.02	60.31	60.63	63.69	82.73	80.28	82.07	81.71
Qwen-Omni-Turbo	76.08	78.83	60.05	71.82	74.88	66.29	60.75	67.34	54.35	47.59	56.73	52.91	82.83	80.71	84.48	82.72
Open-source Models																
Step-Audio-Chat	85.50	70.20	59.55	71.86	79.23	67.80	59.87	69.01	62.81	55.64	56.23	58.25	81.39	79.53	84.04	81.73
GLM-4- Voice	72.41	75.45	64.33	70.81	79.10	74.87	63.02	72.35	57.09	57.43	57.48	57.33	82.04	80.46	82.94	81.86
VITA-Audio-Plus- Vanilla	79.56	67.50	59.22	68.70	79.23	72.73	63.77	71.94	59.45	55.89	57.11	57.50	80.81	80.54	85.20	82.20
MiniCPM-o 2.6	63.15	57.11	46.04	55.74	71.64	70.45	58.11	66.75	49.63	51.07	49.69	50.13	75.28	75.67	78.51	76.50
Kimi-Audio	65.56	50.10	50.00	55.17	74.86	75.63	61.73	70.66	55.33	60.62	58.63	58.17	78.59	79.49	76.01	77.98
Moshi	35.05	46.58	29.58	37.26	46.39	47.15	37.74	43.76	29.73	32.15	28.68	30.18	61.62	58.99	60.93	60.55
AnyGPT	28.57	37.39	32.47	32.76	56.38	64.90	43.90	55.05	15.30	11.94	7.04	11.44	37.20	38.87	44.03	40.06
Human	63.90	61.49	63.55	62.97	71.14	66.67	63.02	66.96	45.02	50.06	58.36	51.13	66.67	71.58	71.82	69.96