Task	Model	Metric	Public Test Set	Public Result	Internal Test Set	Internal Result
SRWT	Whisper-L-v3	AAS(ms, ↓) (Shi et al., 2022)	The comparison plan test set is not open source.	—	$\text{Test}_{\text{srwt}}$	11.09
	Qwen-Audio			—		9.17
	GMM-HMM			—		**7.55**
	OSUM			—		7.61
VED	Qwen2-Audio	ACC (%, ↑)	**ESC-50** **VocalSound**	- 93.3	Test_{ved}	33.25
	TouchASP			85.7 -		—
	PANNs			83.3 -		3.25
	OSUM			96.02 80.75		**81.00**
SER	Qwen2-Audio	ACC (%, ↑)	**MELD** test **MER2023** test	55.3 -	Test_{ser}	38.04
	TouchASP			50.5 -		—
	Sensevoice-L			**63.1** 69.2		—
	Sensevoice-S			57.8 68.3		40.77
	Emotion2Vec			51.88 —		51.14
	EmoBox			51.89 65.23		**74.54**
	OSUM			56.64 **88.84**		72.97
SSR	GLM-4	ACC (%, ↑)	The comparison plan test set is not open source.	—	Test_{ssr}	53.97
	OSUM			—		**68.56**
SGC	Qwen2-Audio	ACC (%, ↑)	**AISHELL-1** test **Kaggle-CommonVoice** valid-test	97.36 97.25	Test_{sgc}	**98.43**
	OSUM			**100** **98.75**		93.74
SAP	Qwen2-Audio	ACC (%, ↑)	**Kaggle-CommonVoice** valid-test	35.53	Test_{sap}	49.52
	OSUM			**76.72**		**67.94**
STTC	Qwen2-Audio	GPT-3.5-Turbo Scoring	**AirBench** speech	6.77	$\text{Test}_{\text{sttc}}$	**6.91**
	ASLP-Audio			4.96		6.69