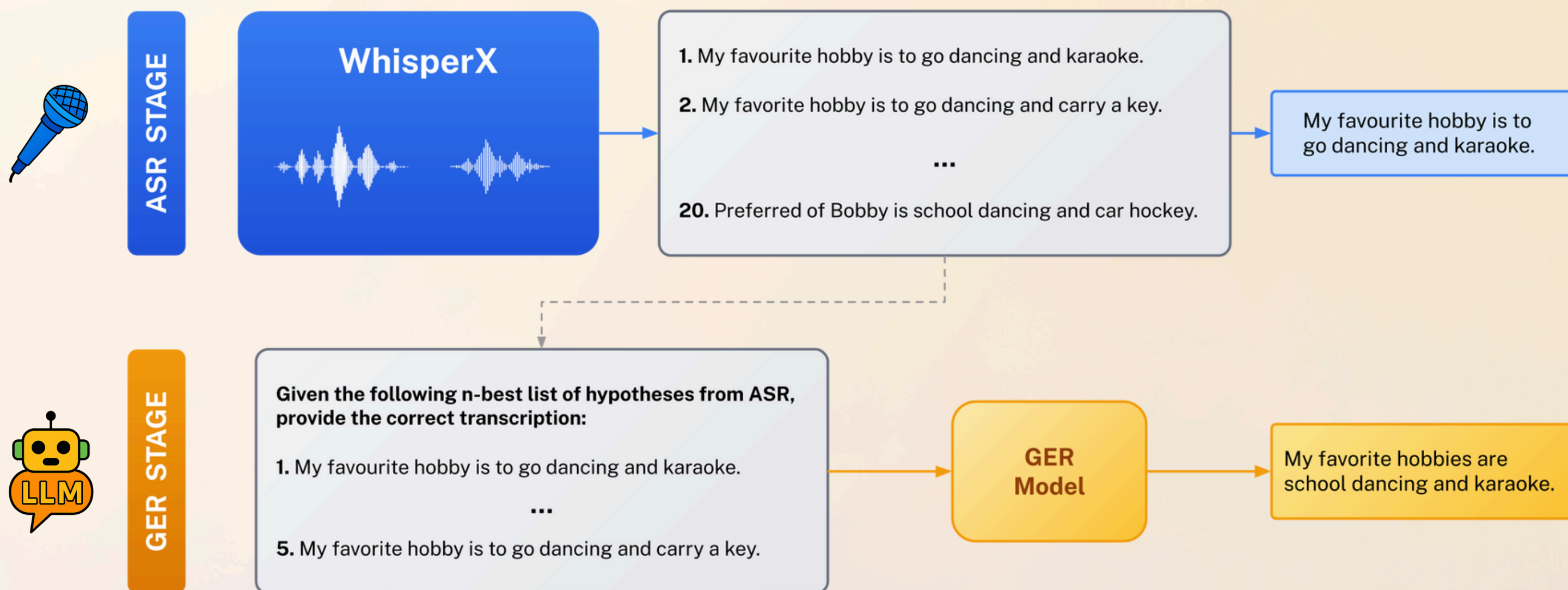


Exploring Generative Error Correction for Dysarthric Speech Recognition



Two-Stage Framework

ASR: Whisper generates 20-best hypotheses

Diversity: picks 5 different hypotheses

GER: FlanT5 model generate final output

Does it work?

(ZS) 11.60% → 7.34% (FT) 7.17% → 6.40%

🏆 Best on Digital Assistant and Sentences

👎 Worst on isolated words.

ASR Model	Fine-Tuned	GER	Dev		Test-1		Test-2	
			WER ↓	SemScore ↑	WER ↓	SemScore ↑	WER ↓	SemScore ↑
Large-v3	×	×	11.60	83.91	11.39	84.82	14.49	78.83
Large-v3	×	3B	7.91	89.42	10.99	86.45	13.87	79.88
Large-v3	×	11B	7.34	90.32	10.63	86.96	13.64	80.40
Large-v2	✓	×	7.17	91.94	10.90	87.24	13.04	81.89
Large-v2	✓	3B	6.40	92.47	10.67	87.51	12.89	82.16

Category	Count	Without GER		With GER	
		WER ↓	SemScore ↑	WER ↓	SemScore ↑
Digital Assistant	15,066	6.47	92.77	5.61 (-0.86)	93.28 (+0.51)
Sentences from Novels	3,746	5.74	92.96	5.21 (-0.53)	93.41 (+0.45)
Spontaneous Speech	2,313	10.86	87.27	10.31 (-0.55)	88.14 (+0.87)
Single Words	130	63.08	49.84	63.08 (+0.00)	49.46 (-0.38)



Repository