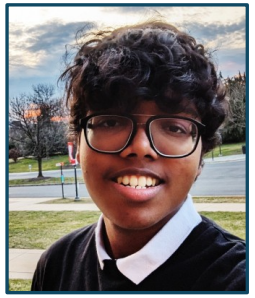
SING: Spatial Context in Large Language Model for Next-Gen Wearables

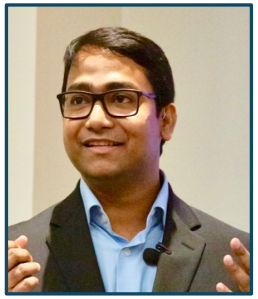
Ayushi Mishra (co-primary)

Yang Bai (co-primary)



Priyadarshan Narayanasamy

Nakul Garg

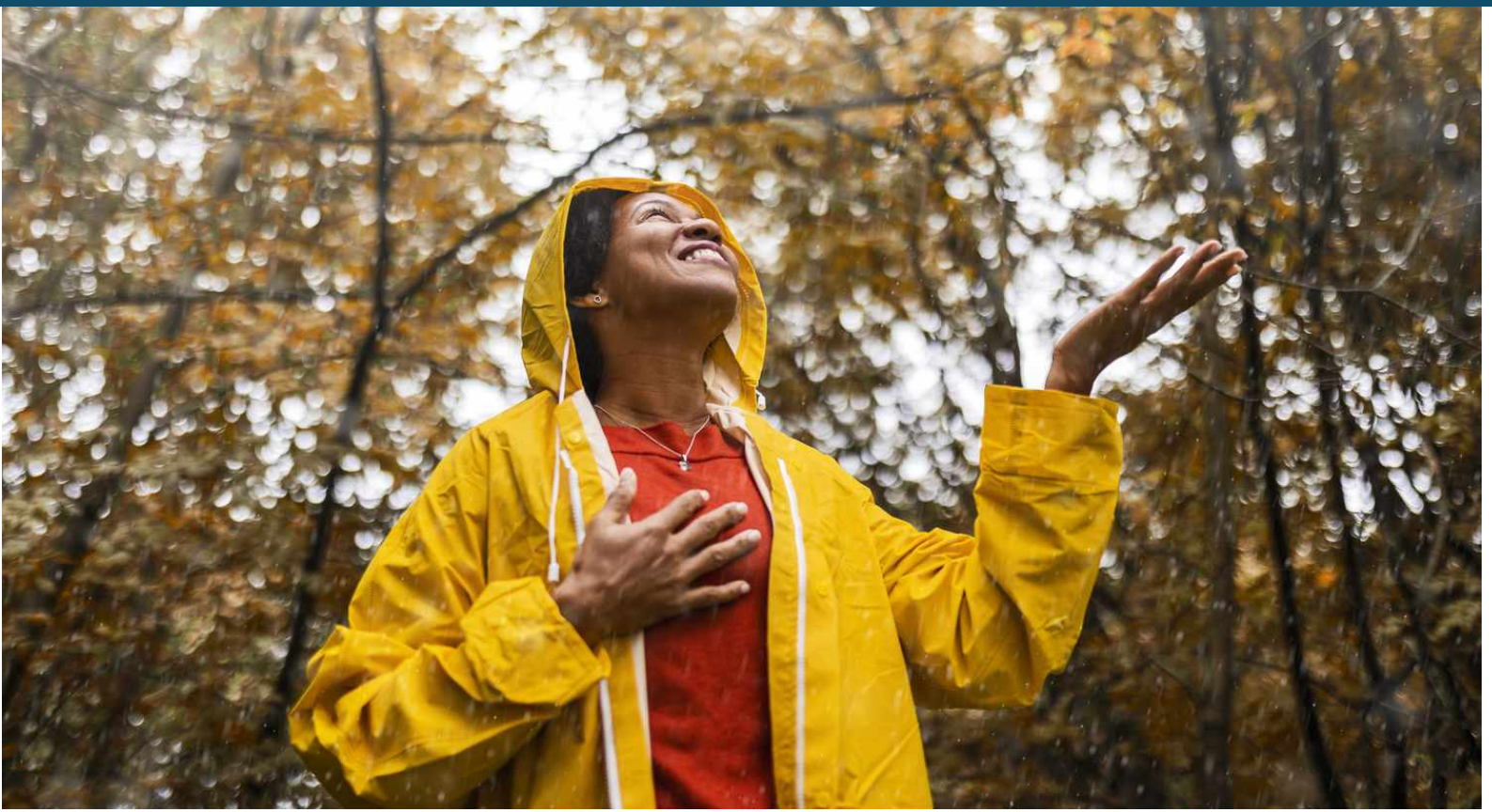


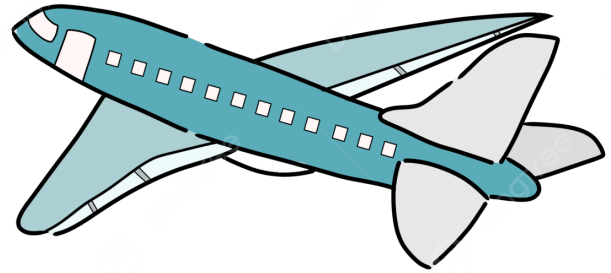
Nirupam Roy

ICML International Conference On Machine Learning July 2025

iCoSMoS Intelligent • Connected • Secure • Mobile • Systems

A World Alive with Sound





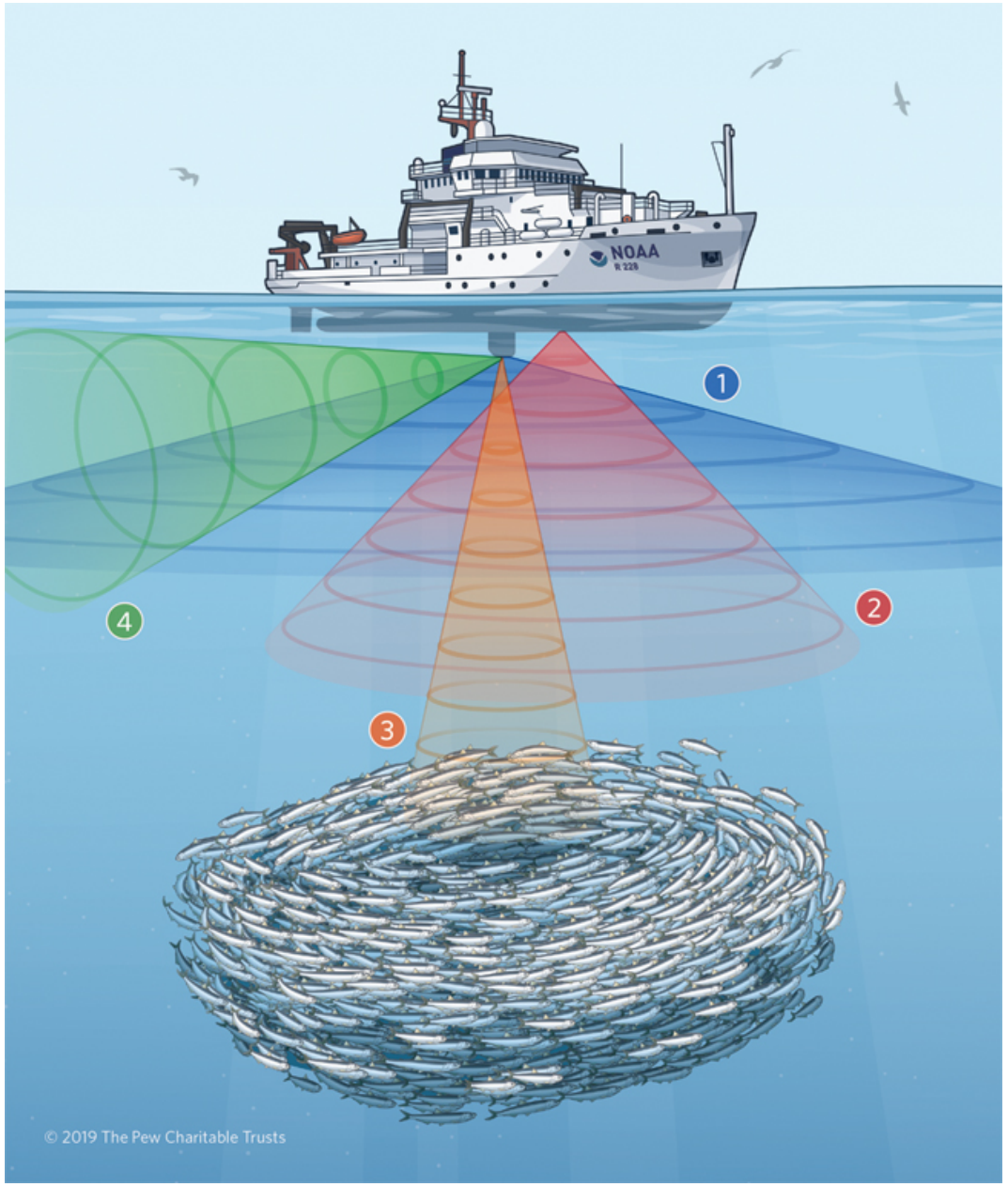
alexa

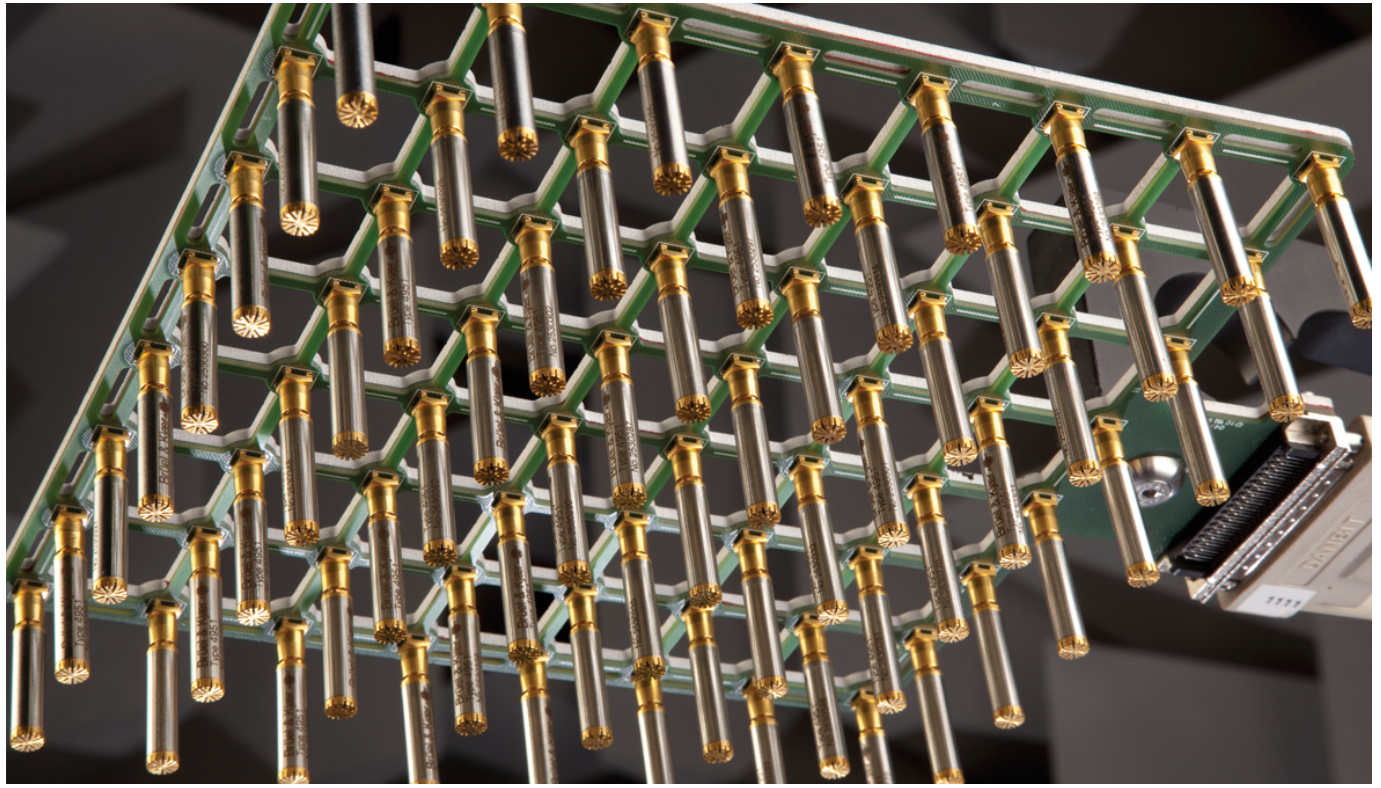
Google Assistant

AirPods

Sound Becomes Smart

Spatial perception with acoustics

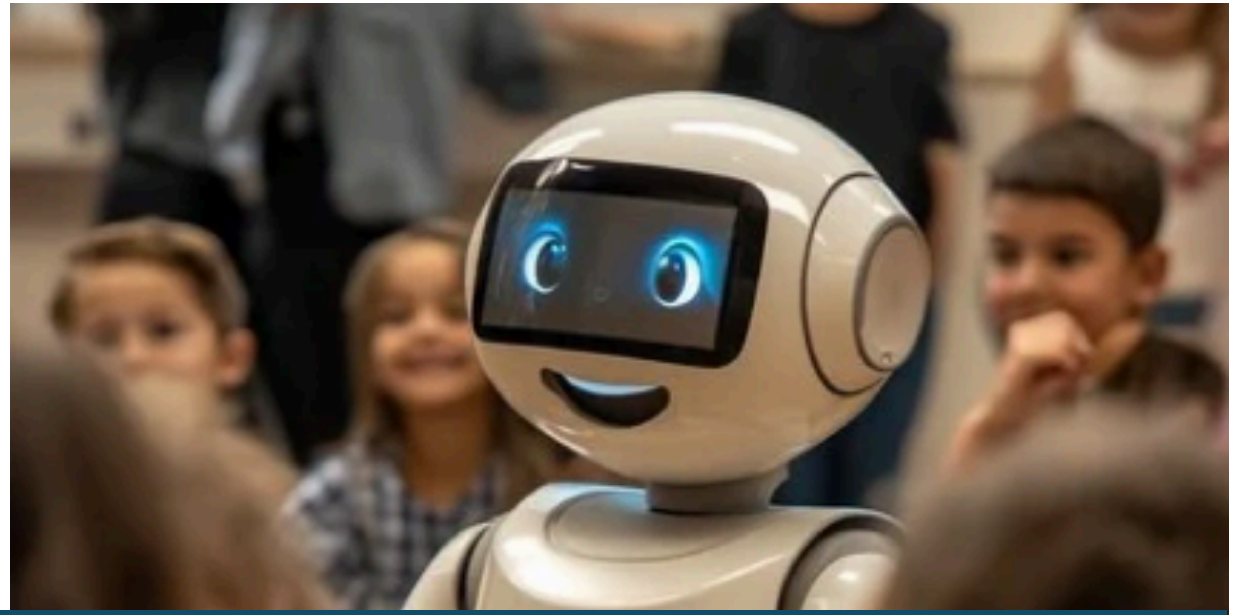




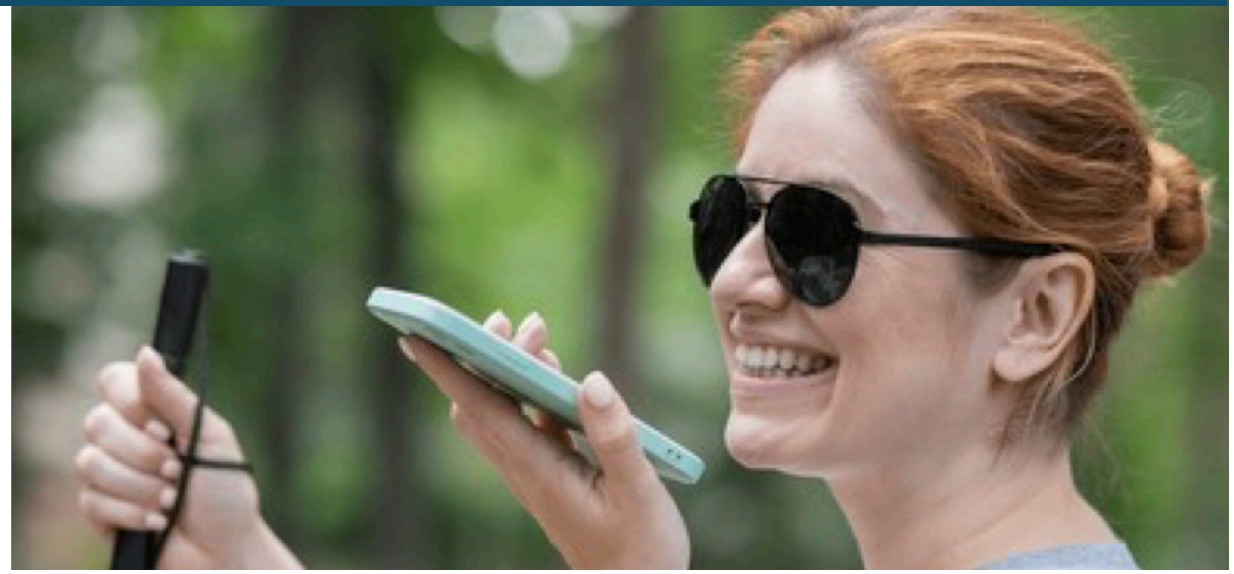
Sonar and microphone array system:
For spatial analysis of sound

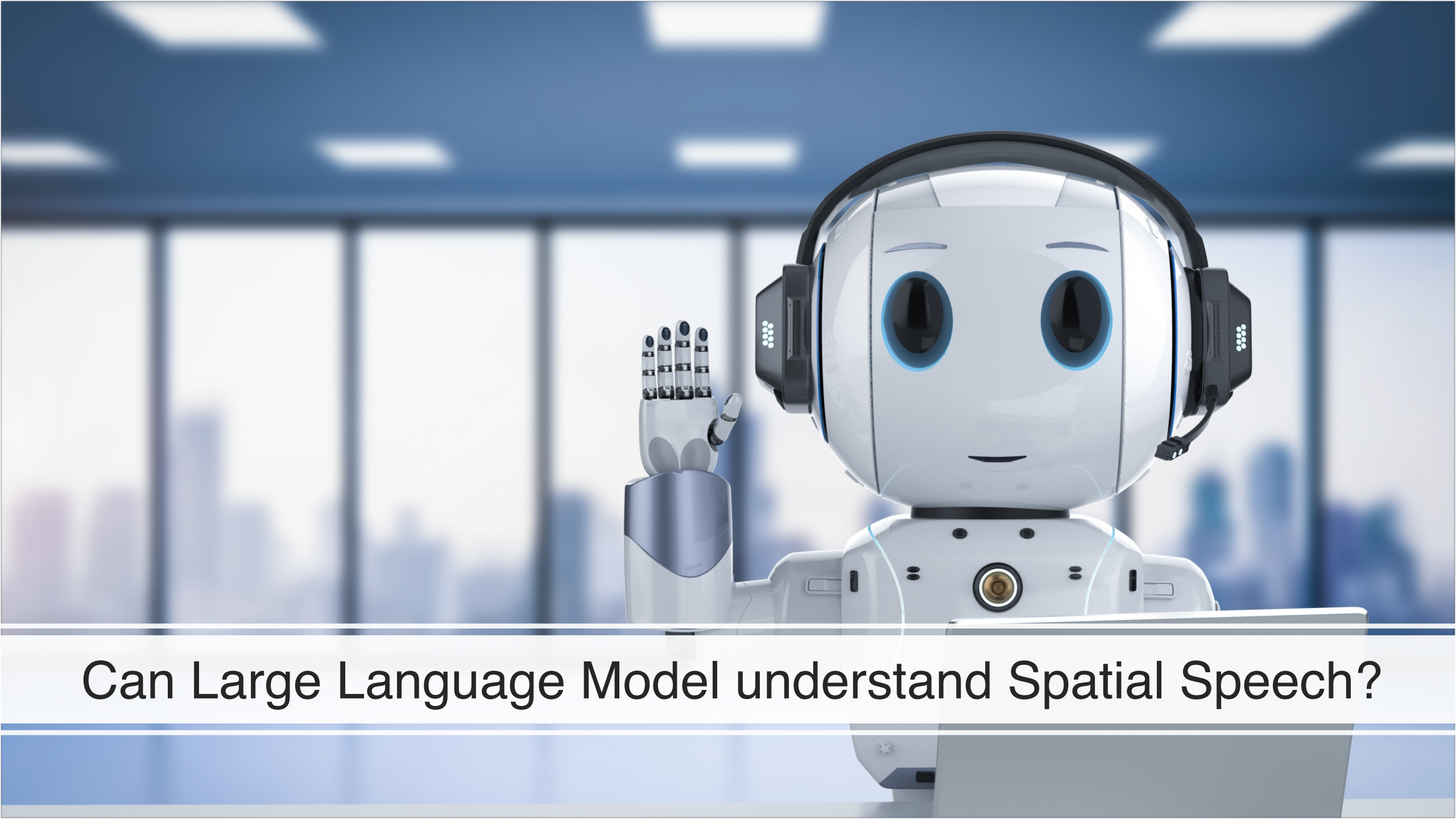
Spatial acoustic sensing in nature

Animals use body structure for spatial audio.



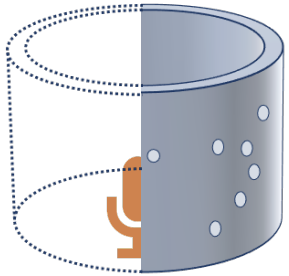
Spatial Soundscape





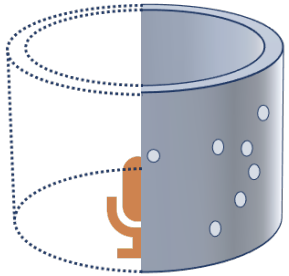
Can Large Language Model understand Spatial Speech?

Core idea: Structure-assisted spatial sensing

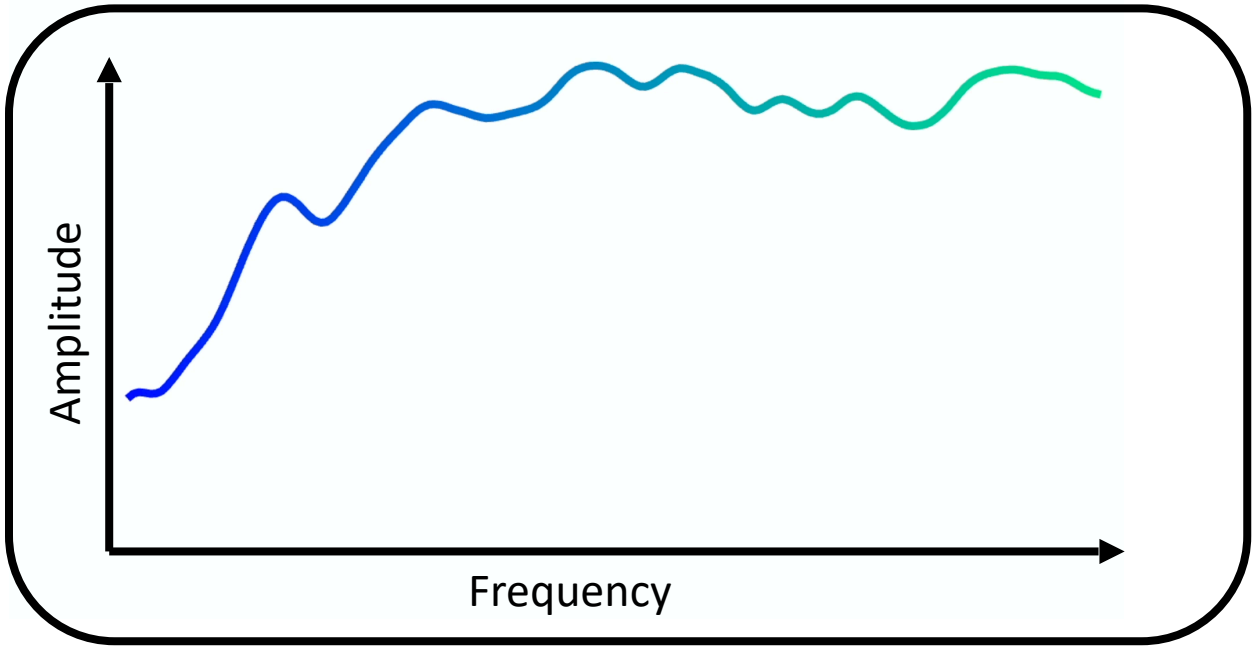


Metamaterial

Core idea: Structure-assisted spatial sensing

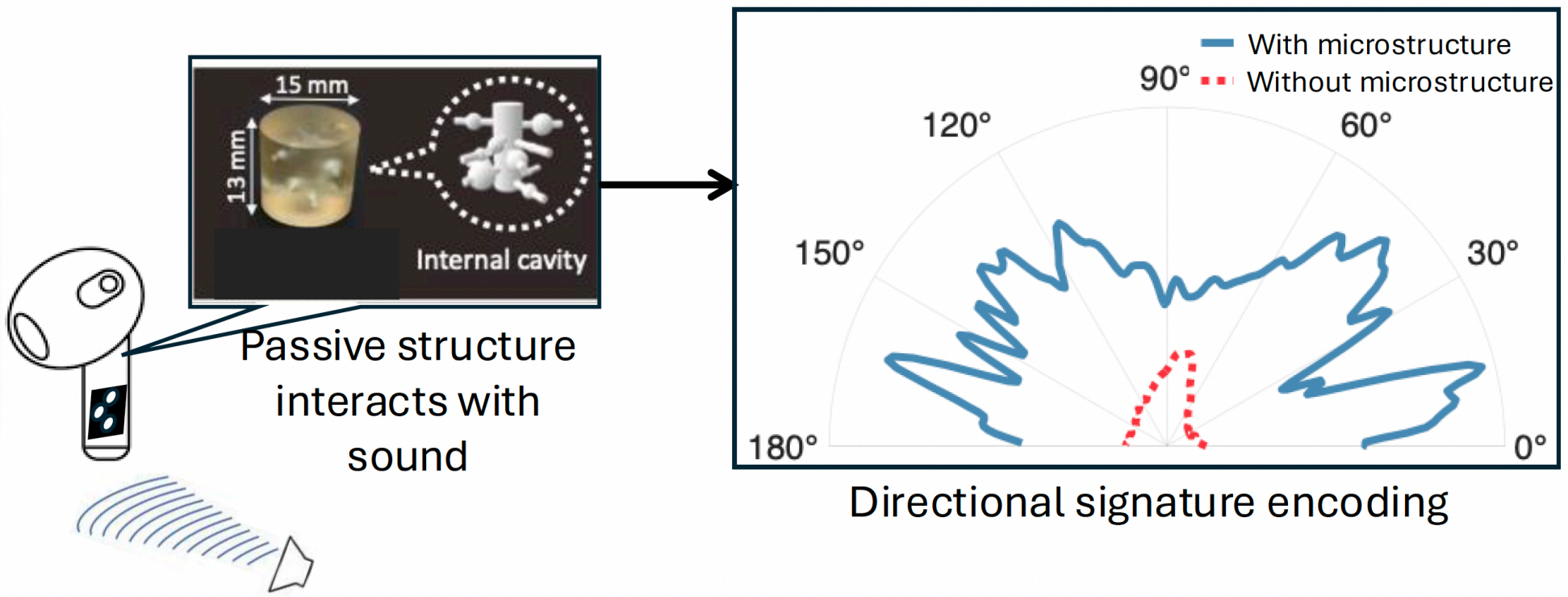


Metamaterial



Received sound signature

Spatial Speech

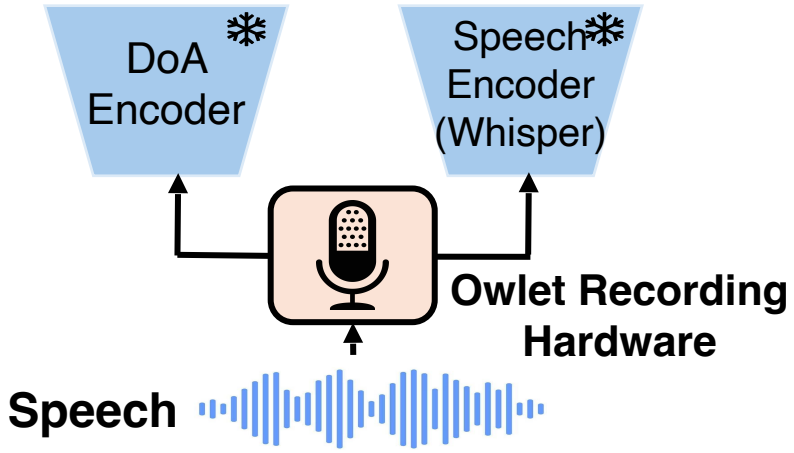


Passive structure interacts with sound

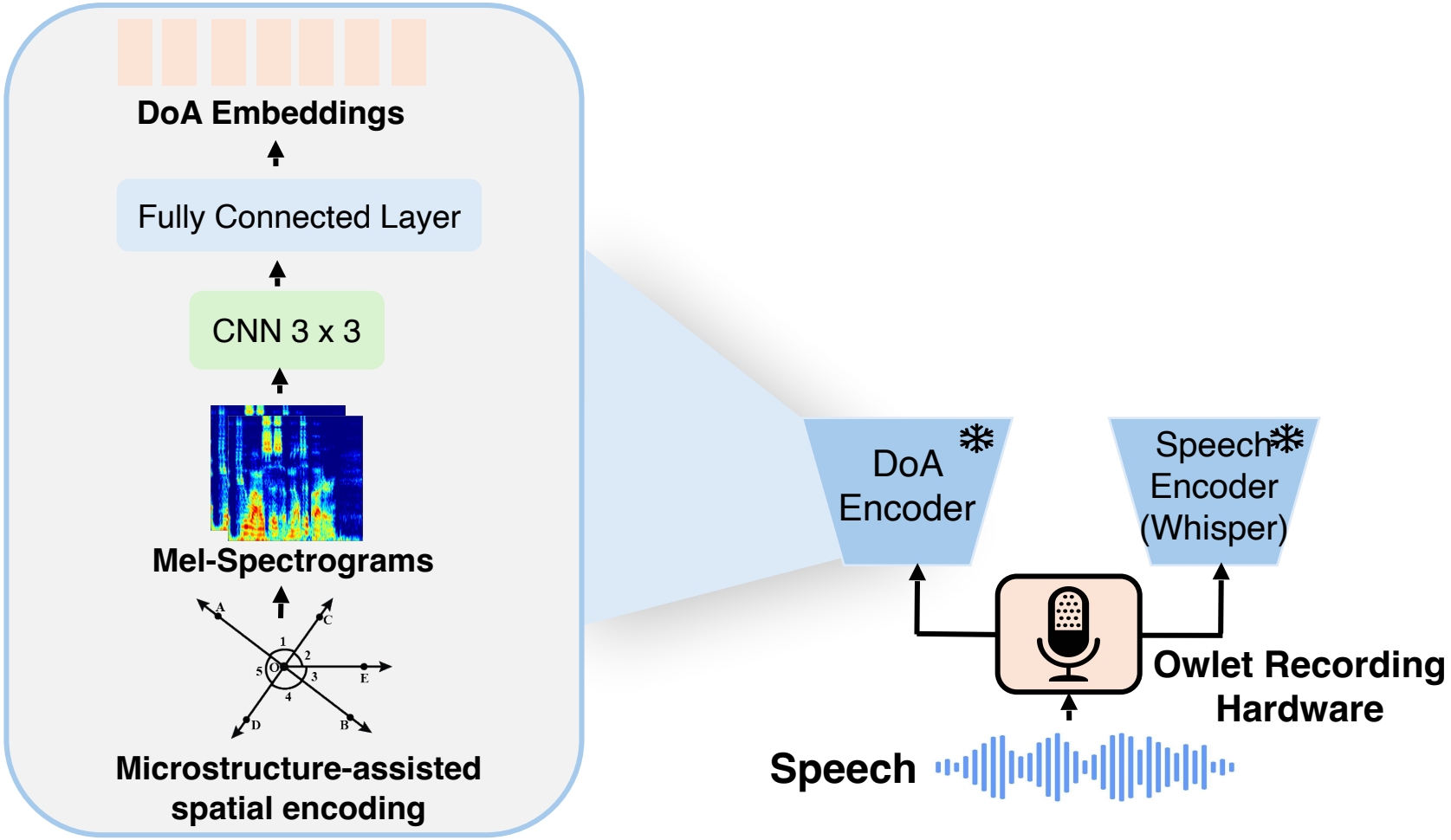
Directional signature encoding

Low-power and miniaturized spatial sensing with acoustic micro-structures

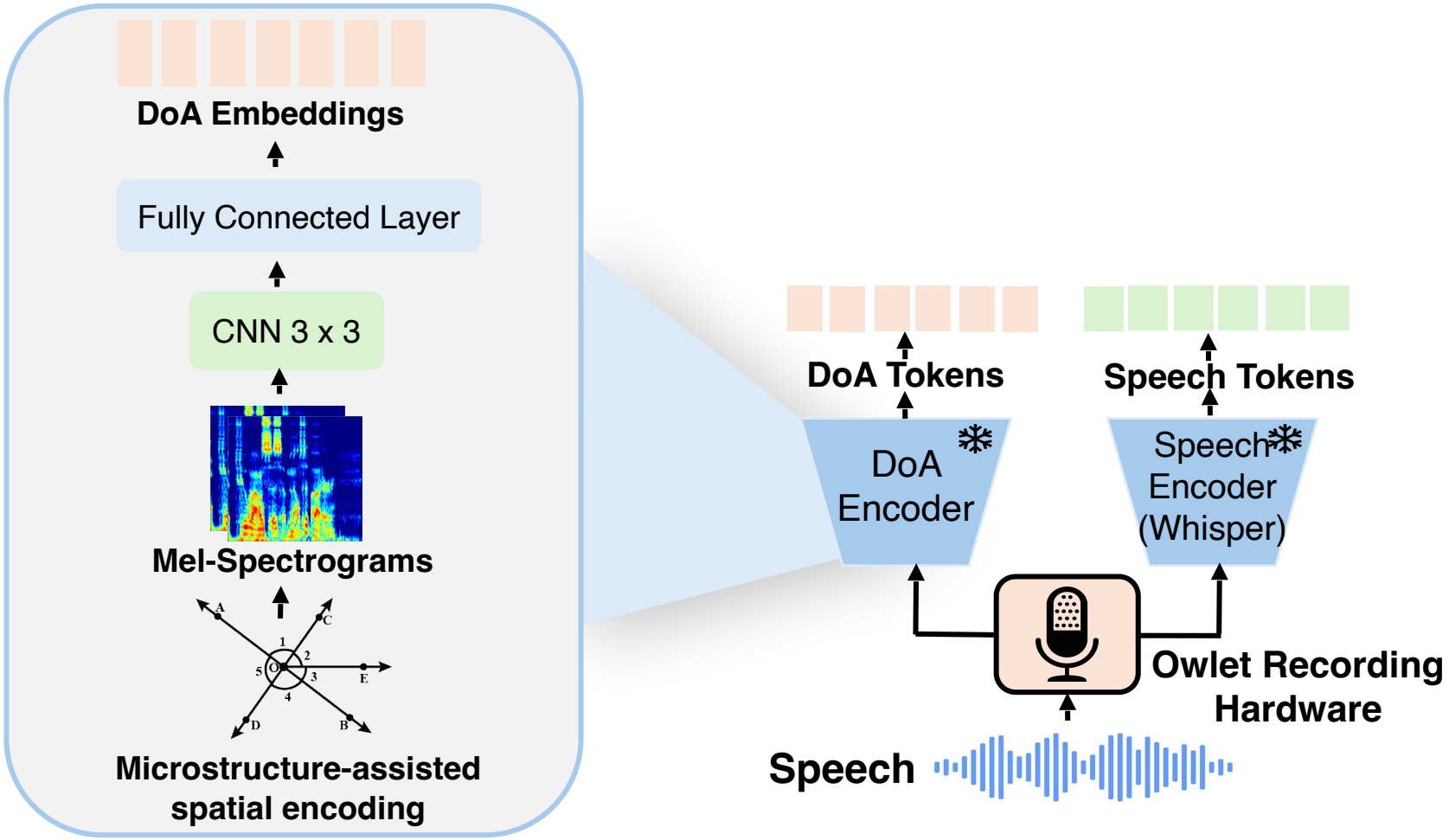
Extract speech feature



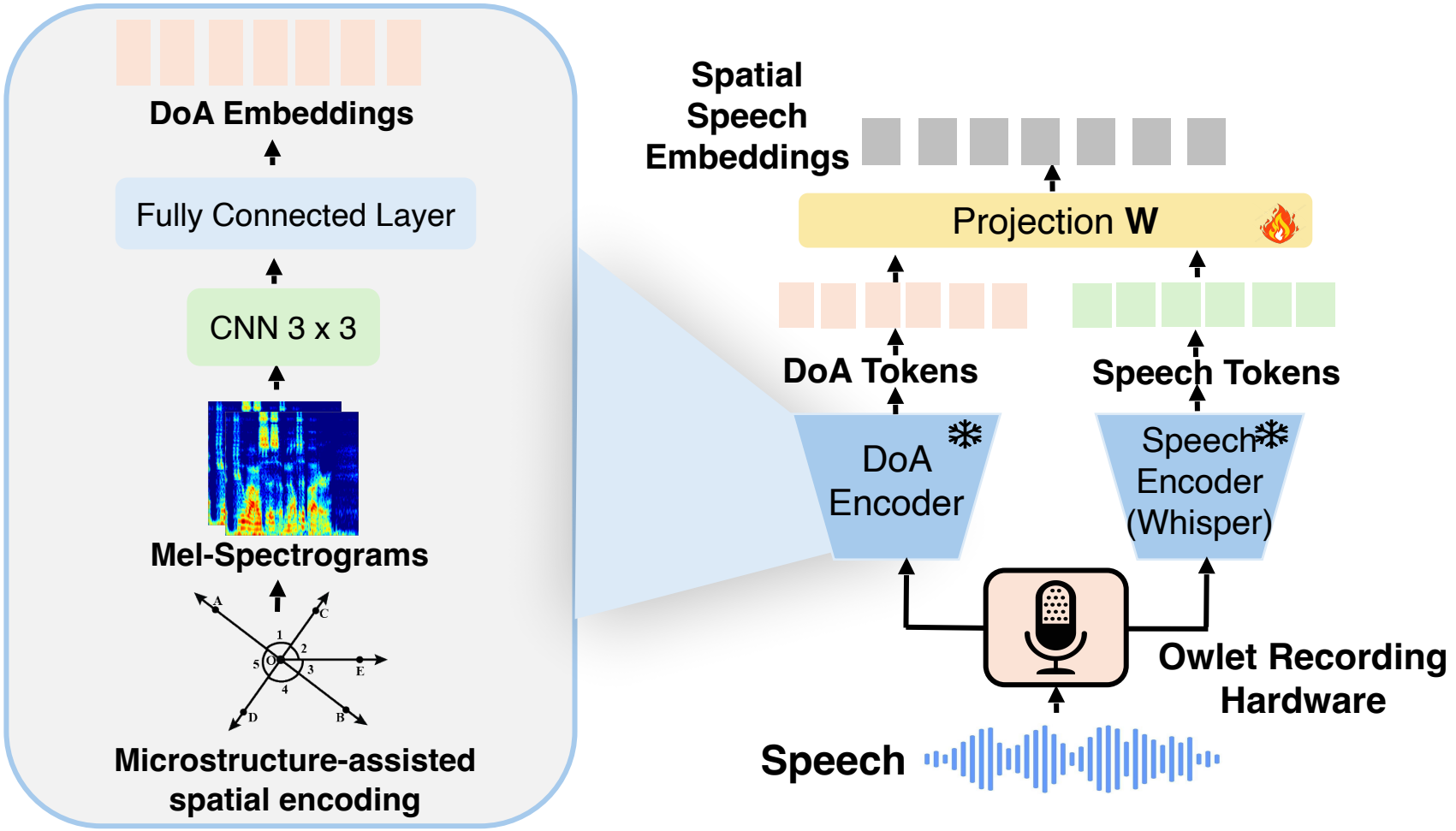
Extract speech feature



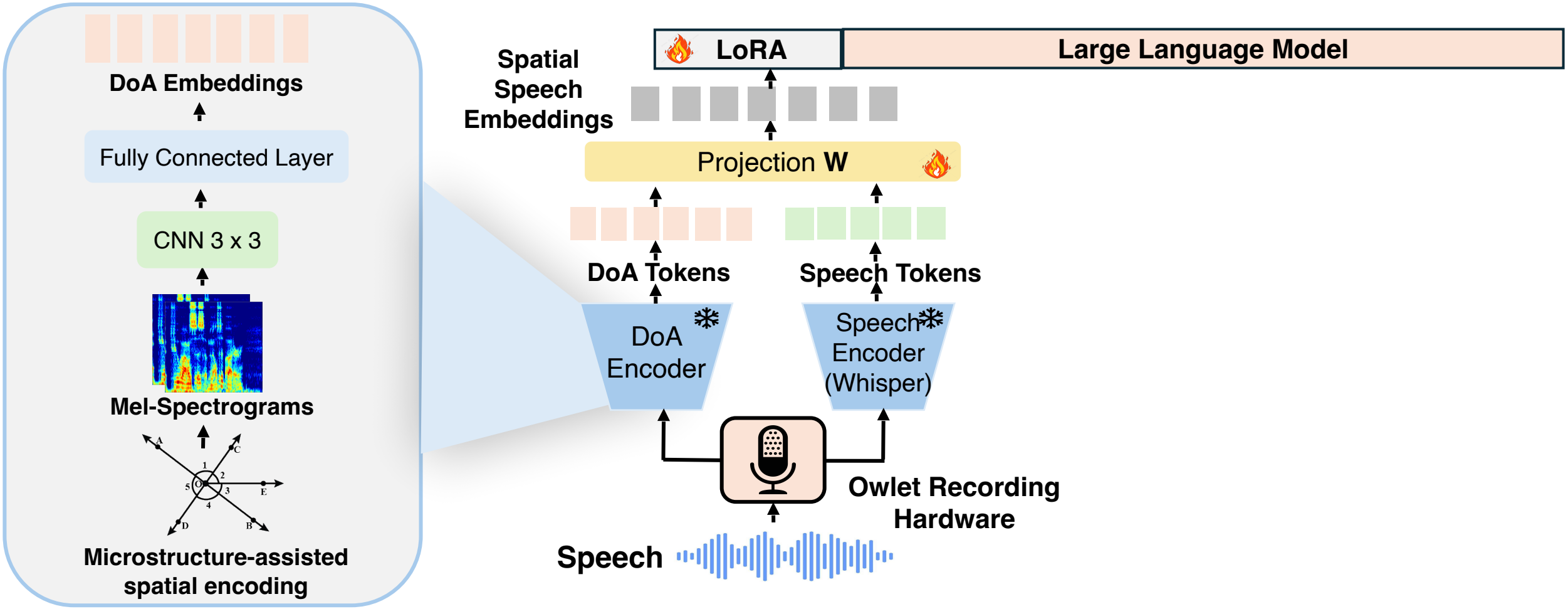
Extract speech feature



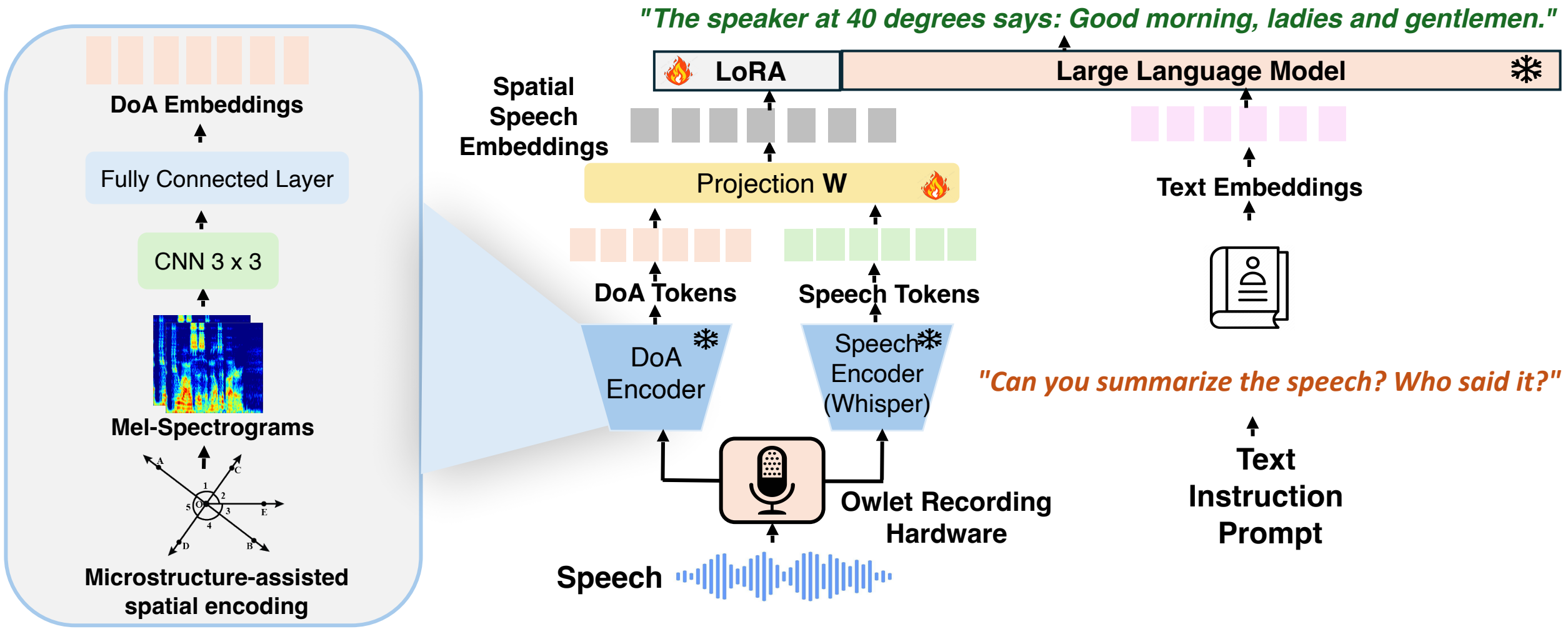
Alignment with LLM space



Fine-tuning LLM for spatial speech



Fine-tuning LLM for spatial speech



Model	Supported Task	Metric / Performance
BAT	DoA Estimation + Audio Source Recognition (No Speech)	MAE (°) ↓: 88.52
SING [our work]	DoA + Speech ASR (Spatial Awareness)	MAE (°) ↓: **25.72**
SALMONN	Speech ASR and LLM Functions (No Spatial Awareness)	WER (%) ↓: 2.2
SING [our work]	Speech ASR (without DoA)	WER (%) ↓: **1.8**
SING [our work]	DoA + Speech ASR (Spatial Awareness)	WER (%) ↓: 5.3

Model	Metric	1 Source	2 Sources	3 Sources	4 Sources	5 Sources
SELDNet	MAE (↓)	90.03	✗	✗	✗	✗
	MEEM (↓)	360.12	✗	✗	✗	✗
	Median Error (↓)	90.14	✗	✗	✗	✗
AudioMAE	MAE (↓)	43.79	✗	✗	✗	✗
	MEEM (↓)	43.79	✗	✗	✗	✗
	Median Error (↓)	27.79	✗	✗	✗	✗
SING (Ours)	MAE (↓)	**25.72**	**24.16**	**28.11**	**23.31**	**17.08**
	MEEM (↓)	**25.72**	**24.16**	**28.11**	**23.31**	**17.08**
	Median Error (↓)	**13.00**	**13.00**	**20.00**	**18.00**	**13.00**

Thank you!

SING: Spatial Context in Large Language Model for Next-Gen Wearables

http://icosmos.cs.umd.edu/images/2_publication/papers/SING_ICML_2025_icosmos.pdf

Contacts:

Ayushi Mishra
amishr13@umd.edu

Nirupam Roy
niruroy@umd.edu

iCoSMoS
Intelligent • Connected • Secure • Mobile • Systems

http://icosmos.cs.umd.edu