CHI 2025

M²Silent: Enabling Multi-user Silent Speech Interaction Via Multi-directional Speakers in Shared Spaces

Juntao Zhou, Dian Ding, Yijie Li, Yu Lu, Yida Wang, Yongzhao Zhang, Yi-Chao Chen, Guangtao Xue

SJTU

Silent Speech Interface

Silent speech interfaces (SSI) has gained popularity recently in environments such as libraries, cafes, and museums.

Silent Speech Interface

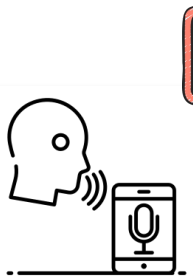
privacy concerns

discomfort from speaking openly

challenges in noisy environments

SSI has various advantages against traditional voice interaction.

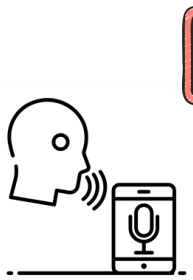
Limitations of Silent Speech Interface



SSI on mobile devices is not sufficient to serve various scenarios.

Occupied hands: Can't use devices when holding a child or items.

Limitations of Silent Speech Interface

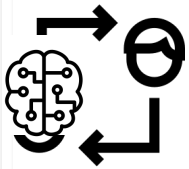


SSI on mobile devices is not sufficient to serve various scenarios.

Special groups: Device use is difficult for the elderly or children.

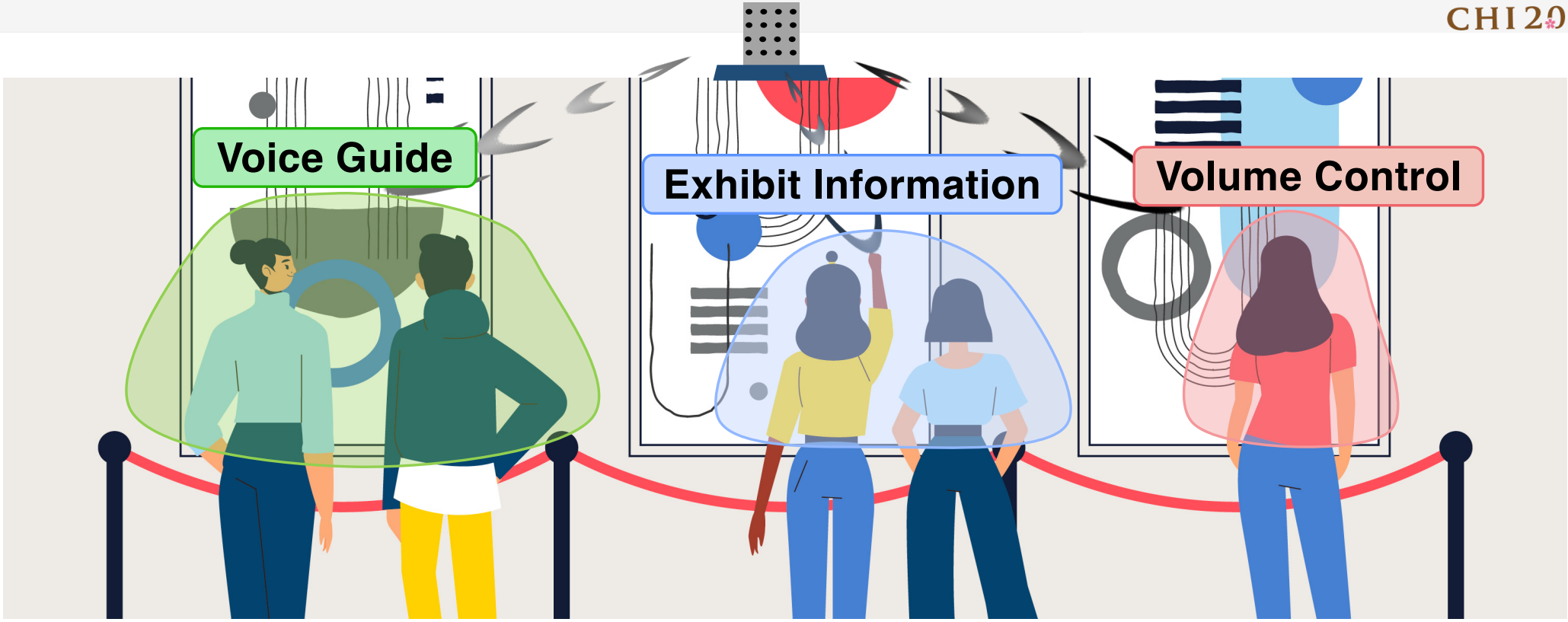
Limitations of Silent Speech Interface

SSI on mobile devices is not sufficient to serve various scenarios.



Bidirectional Interaction: Users not only need to deliver silent commands, but also receive audio feedback.

Multi-user Silent Speech Interaction



In museums, **multi-user silent speech interaction** enables each visitor to have **different inquiries** without **disturbing the environment**.

Multi-user Silent Speech Interaction

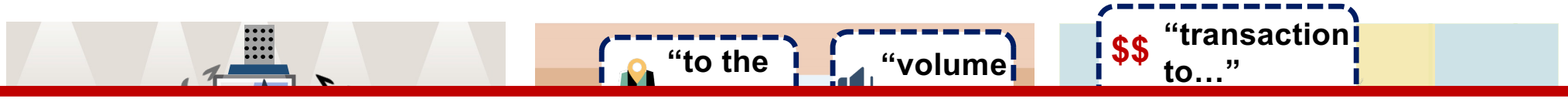
Museum

In-car Space

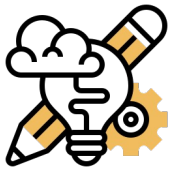
Bank Counter

Multi-user silent speech interaction can work in various open scenarios.

Multi-user Silent Speech Interaction



Design Space:



1) Covering a large infrastructure

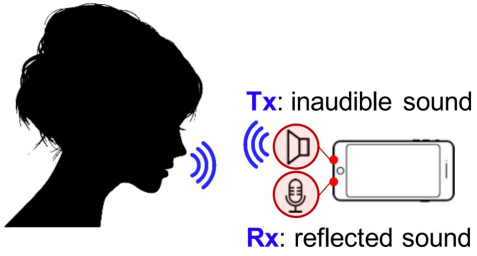
2) Simultaneous use by multiple people

3) Supporting bidirectional interaction

Multi-user silent speech interaction can work in various open scenarios.

Existing Solutions

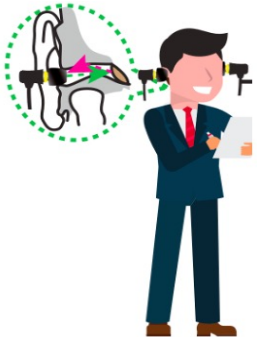
✖ Requiring a personal mobile device



[Endophasia - IMWUT'20]

[Lipwatch - IMWUT'24]

[LipLearner - CHI'23]



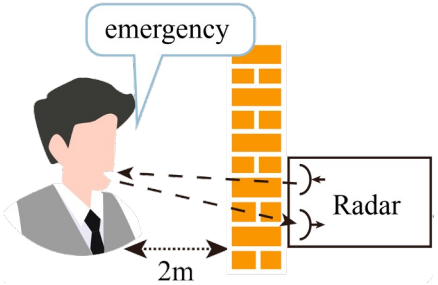
[EarCommand - IMWUT'22]

Existing Solutions

CHI2025

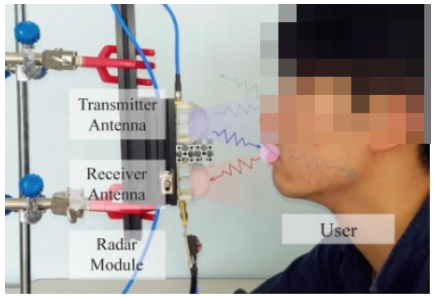
- ~~Requiring a personal mobile device~~
- **Unable to achieve speech recognition and provide acoustic feedback on one device**

[mSilent - IMWUT'23]



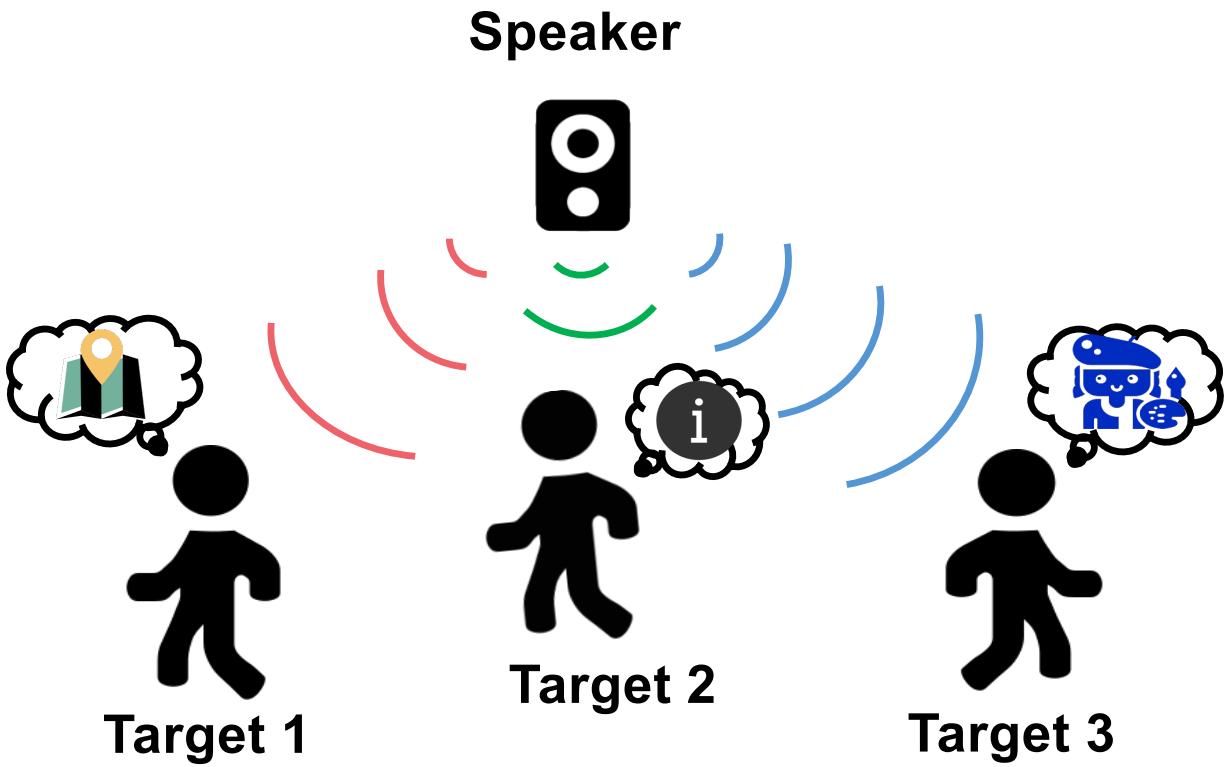
[TWLip - IOTJ'21]

[RaSSpeR - INTERSPEECH'21]

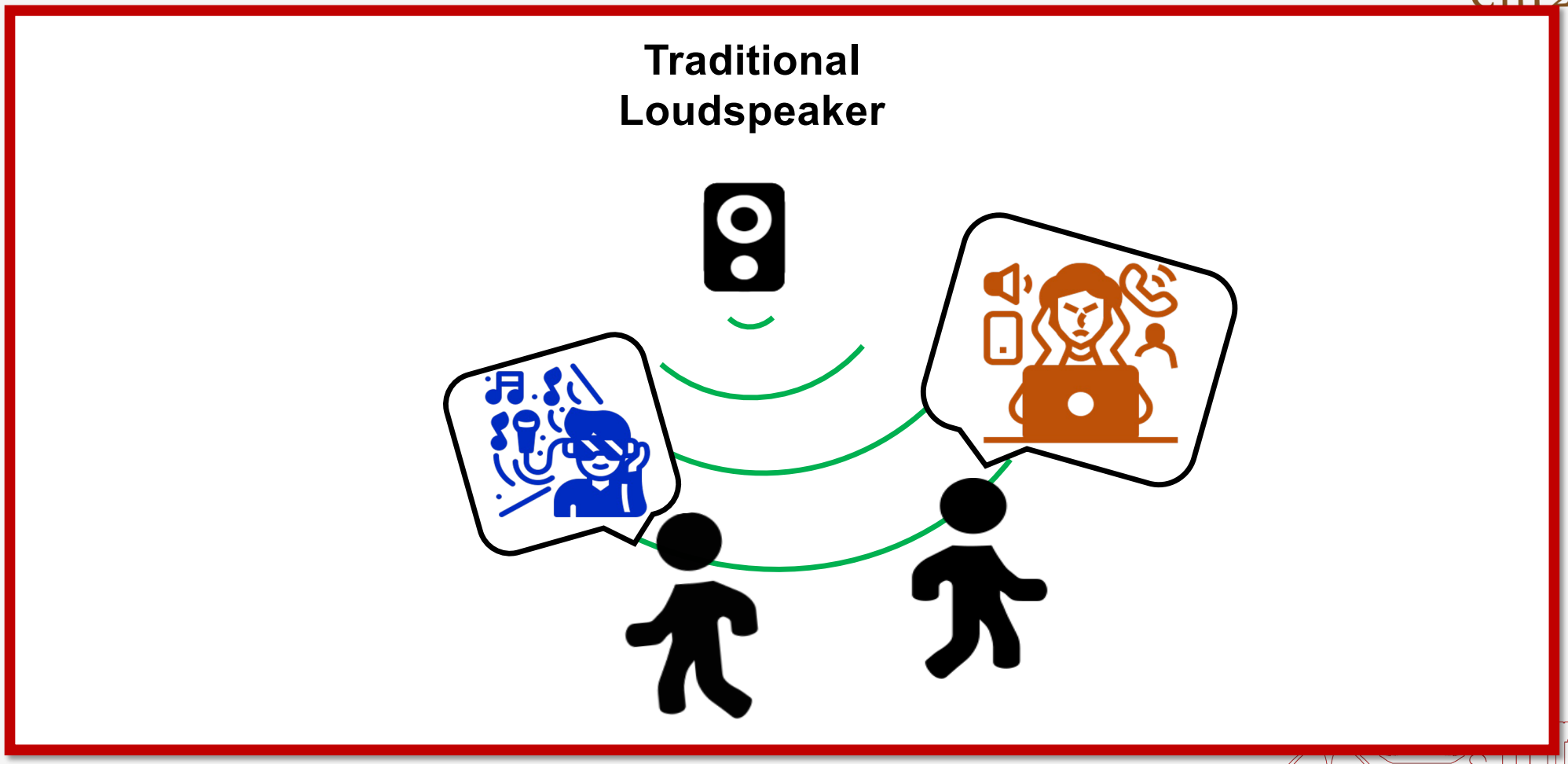


[UWB - Sensors'16]

Our Goal: Multi-user Bidirectional SSI



Our Goal: Multi-user Bidirectional SSI



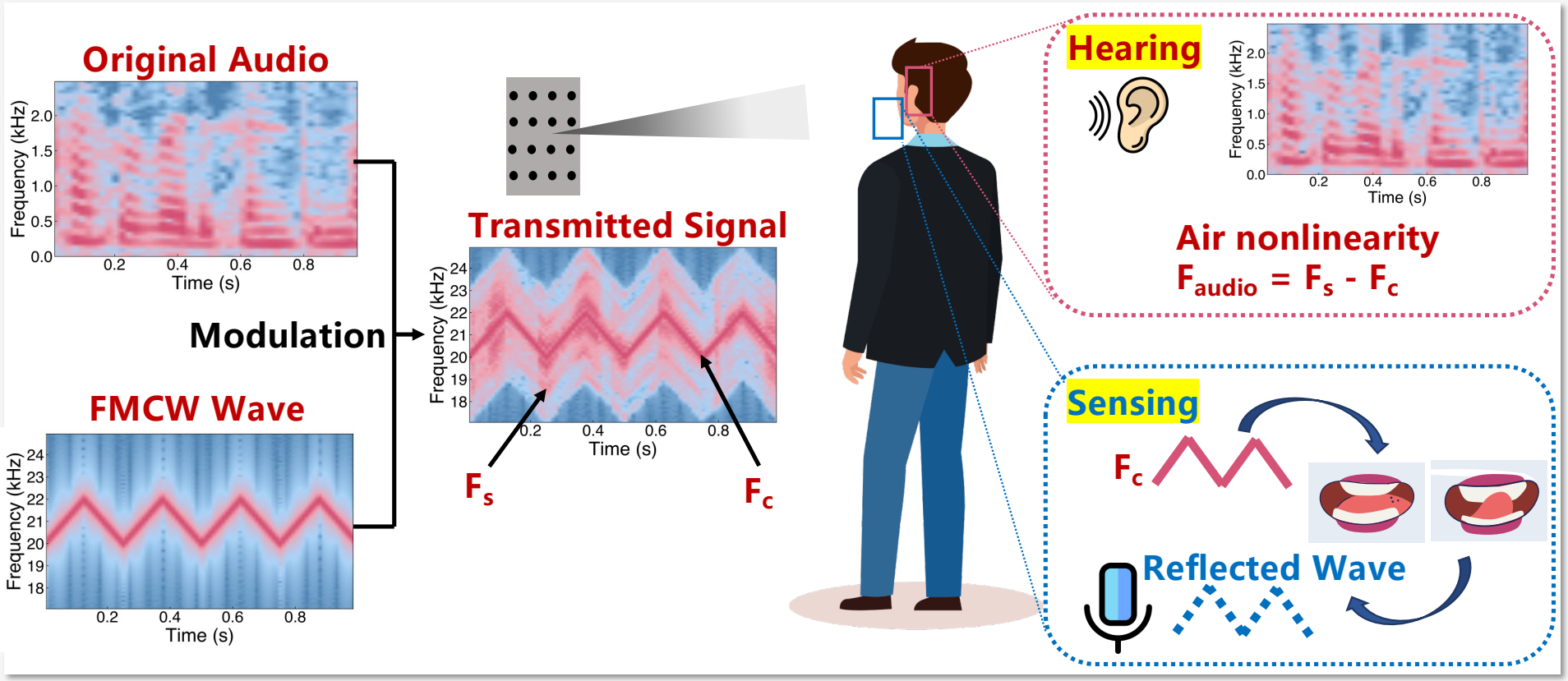
Multi-directional Speaker

[MuDiS - MobiCom'24]

Challenge 1: How can we achieve both audio transmission and silent speech recognition with limited bandwidth?

✖ **Narrow bandwidth limits sensing, causing multi-directional speakers to support only one-way communication.**

Method 1: Empowering Directional Speakers with Sensing

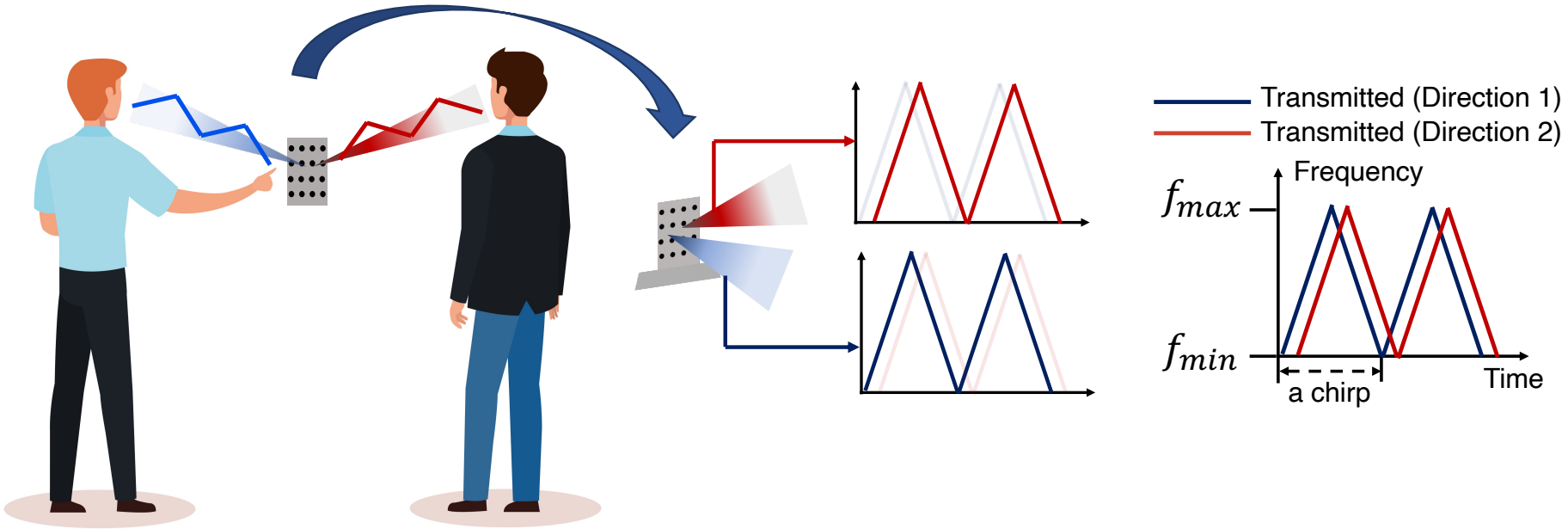


Multi-user Silent Speech Interactions

Challenge 2: How can we distinguish silent speech features from multiple users?

Method 2: Multi-User Silent Speech Feature Segmentation

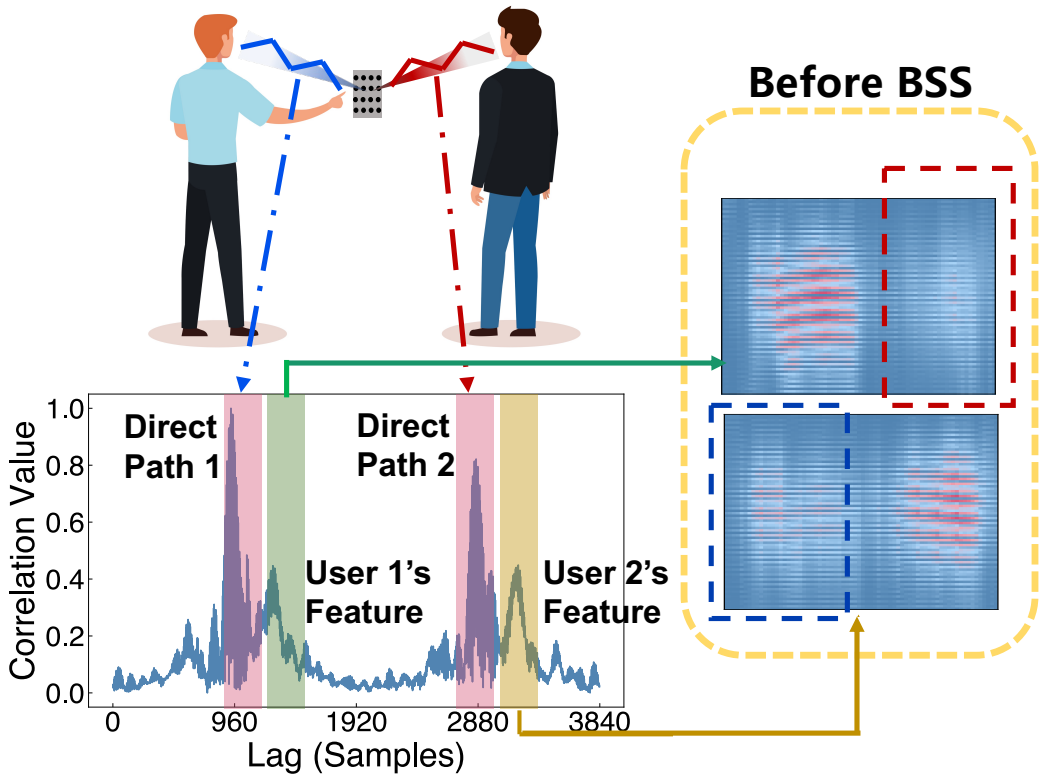
Time-shifted FMCW Signal Emission in Multiple Directions



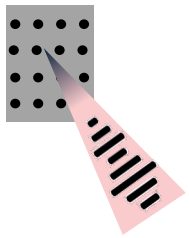
Method 2: Multi-User Silent Speech Feature Segmentation

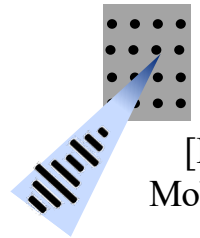
CHI2025

Blind Source Separation of Mixed Features



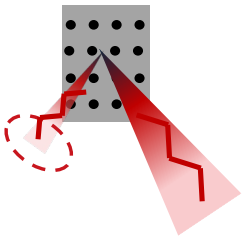
Previous work only eliminated leakage of low-frequency signals.

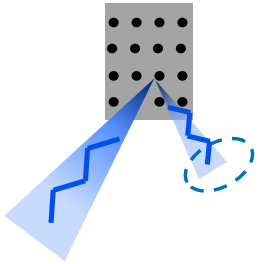




[MuDiS - MobiCom'24]

However, we use high-frequency signals to sense, and there will be leakage.

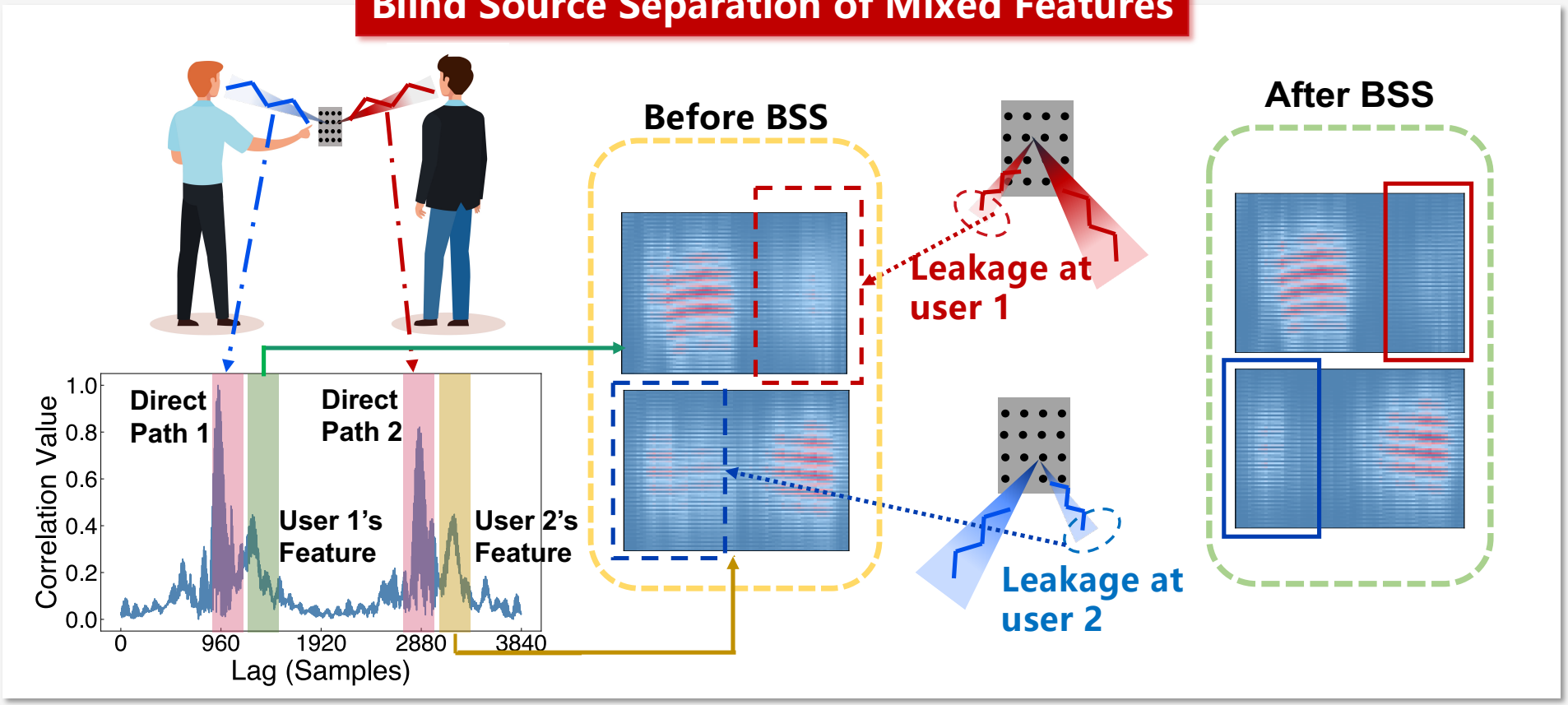




Method 2: Multi-User Silent Speech Feature Segmentation

CHI2025

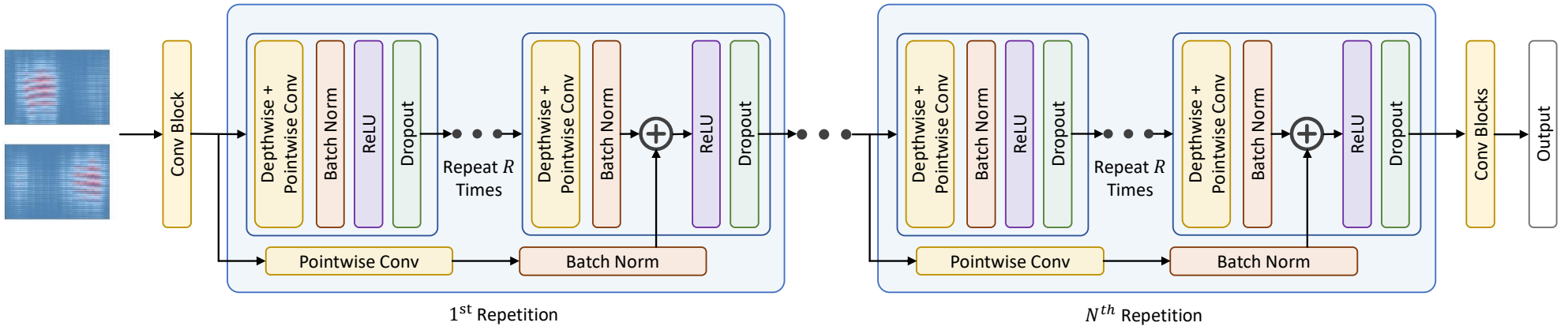
Blind Source Separation of Mixed Features



Silent Speech Recognition

CHI2025

SilentMatch



After we segment the features, we input the clean features into the silent speech recognition model for training and prediction.

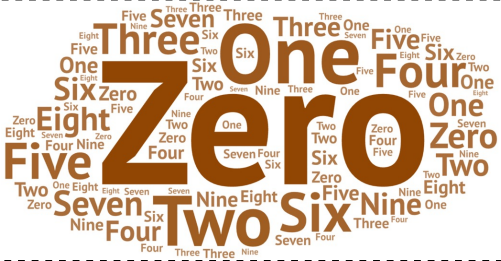
Evaluation

CHI 2025

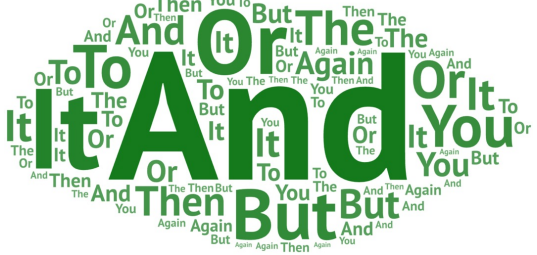
Word Set

Action/ Status

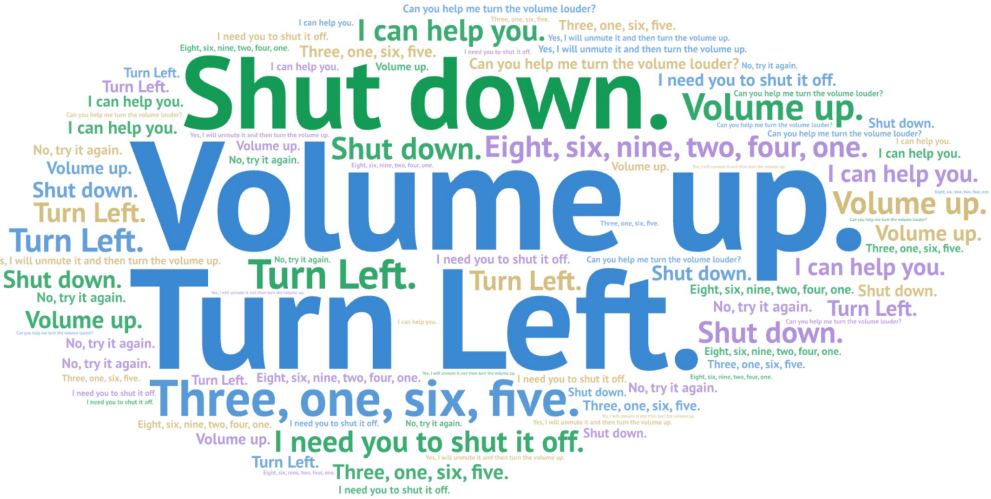
Digits



Conjunctions /Pronouns/...



Possible Sequences

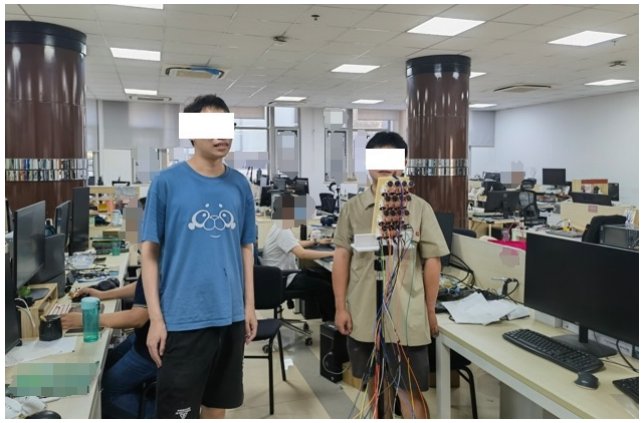


Evaluation

CHI2025

Spacious Indoor

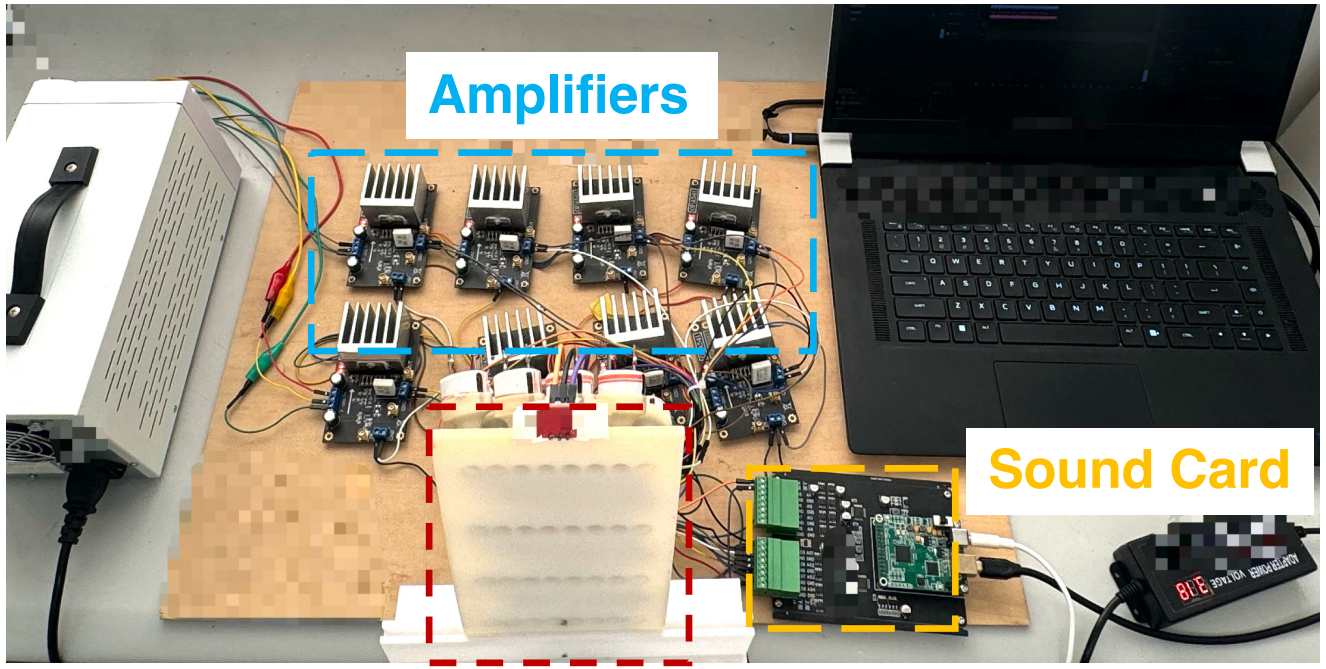
Complex Indoor



Outdoor

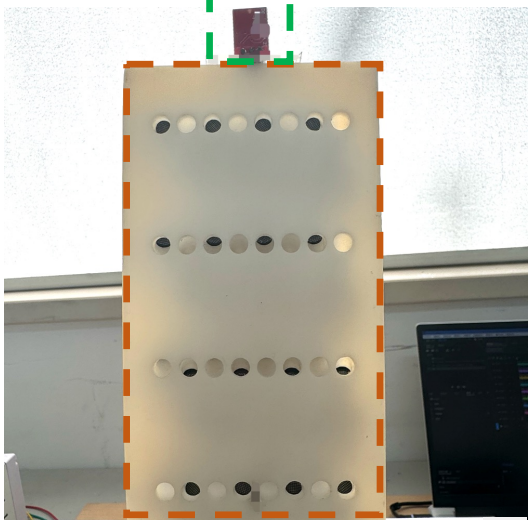
Participants were assigned to 3 different environments to simulate realistic usage scenario.

Evaluation



Multi-directional Speaker

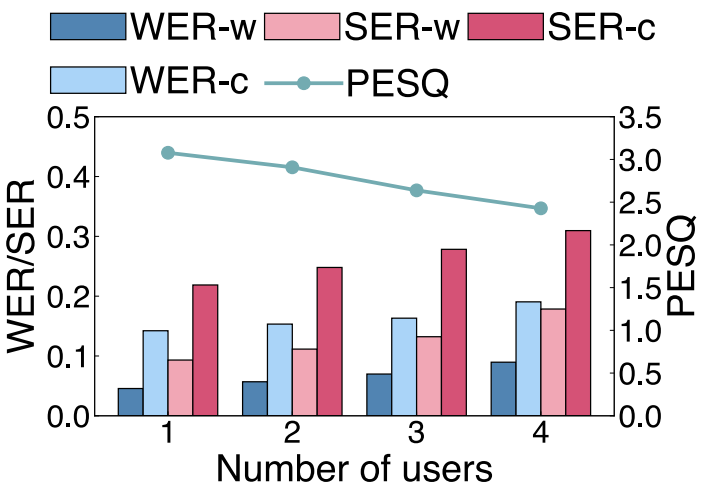
Microphone



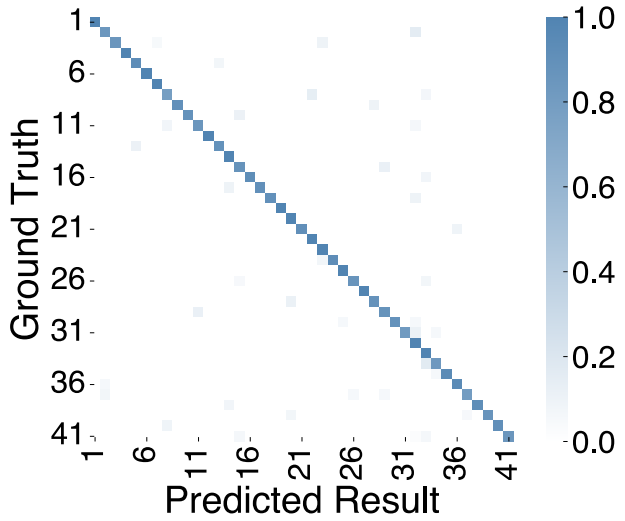
4×8 Transducer Array with Metasurface

Evaluation

Overall Performance

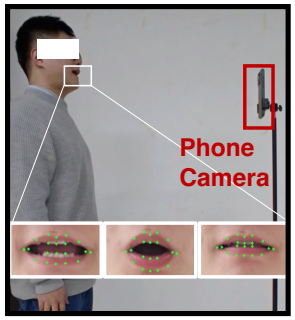
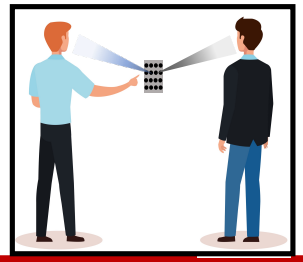


Confusion Matrix



M²Silent achieves WER of 6.5% and SER of 12.8% in multi-user silent speech recognition while maintaining high audio quality.

Evaluation

		WER	Long Distance	Privacy	Dark
Visual Input [Lip-interact - UIST'18]		4.12%	✓	✗	✗
Phone Speaker & Mic [Soundlip - IMWUT'21]		8.26%	✗	✓	✓
M²Silent		6.92%	✓	✓	✓

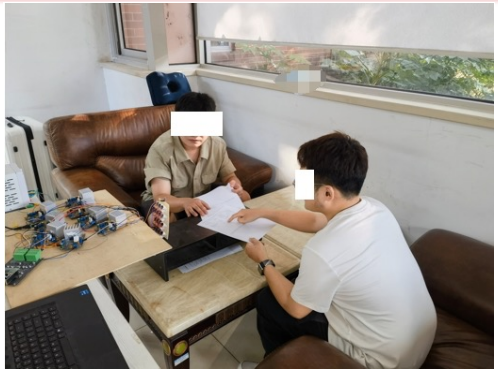
M²Silent can achieve an accuracy comparable to that of traditional silent speech recognition.

Use Case

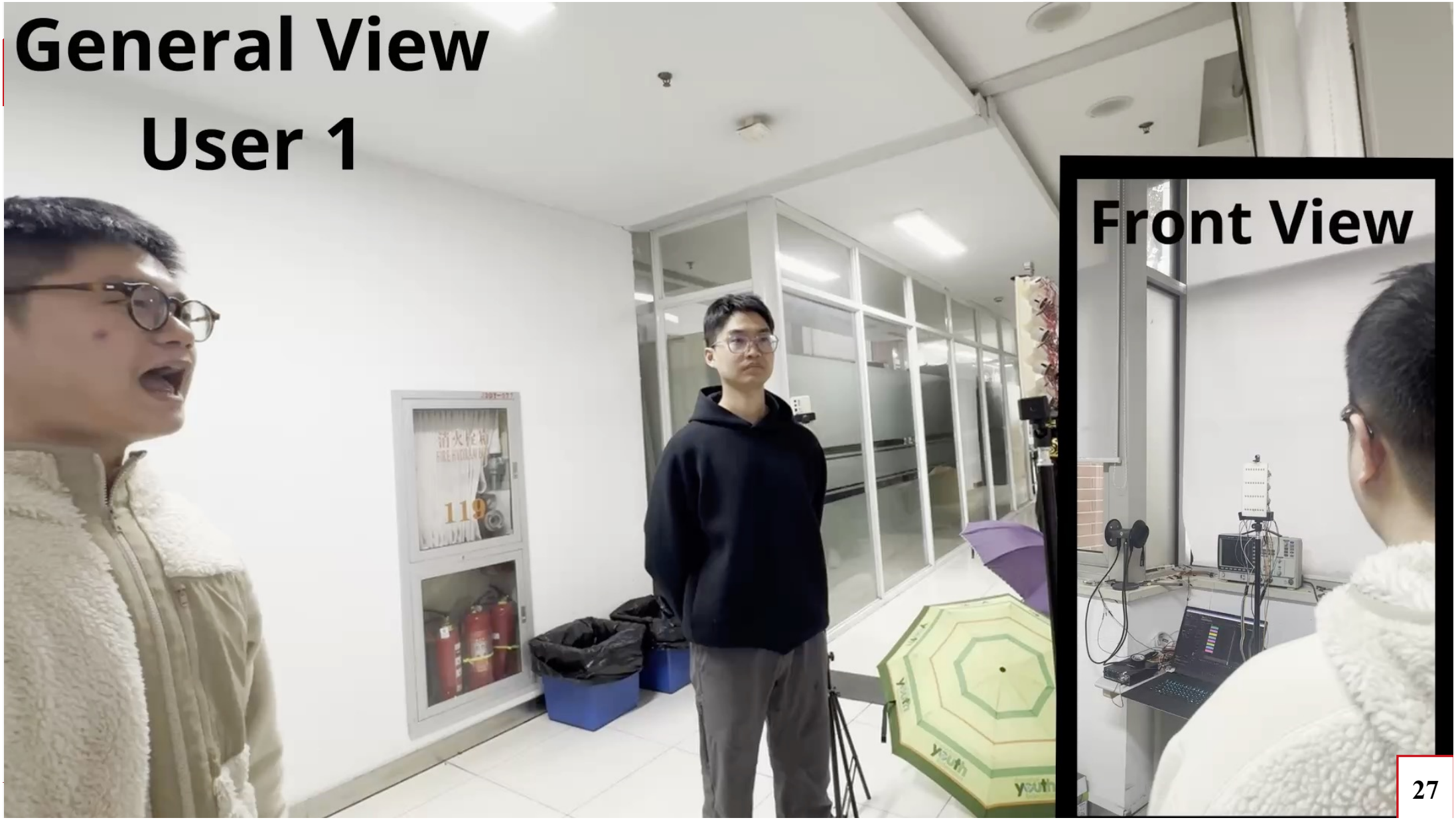
In the exhibition room

In the car

In the transaction



On the street


General View
User 1
Front View

Conclusion

- M²Silent is the first **open-environment silent speech interface** using multi-directional speakers for device-free **multi-user interaction**.
- It employs **FMCW as audio carriers** for simultaneous audio and sensing transmissions.
- It leverages **time-shifted FMCW**, **blind source separation**, and a sliding window approach for multi-user sentence-level recognition.

上海交通大學

SHANGHAI JIAO TONG UNIVERSITY

Thank You

饮 水 思 源　爱 国 荣 校