Face Recognition In Harsh Conditions: An Acoustic Based Approach With Commercial Device

Yanbo Zhang, Panrong Tong, Songfan Li, Yaxiong Xie, Mo Li

MOBISYS '24, Minato-ku, Tokyo, Japan

4 June, 2024

Background

- **The accuracy of vision-based face recognition reduces under harsh environment.**
- **Vision based face recognition causes privacy concern.**

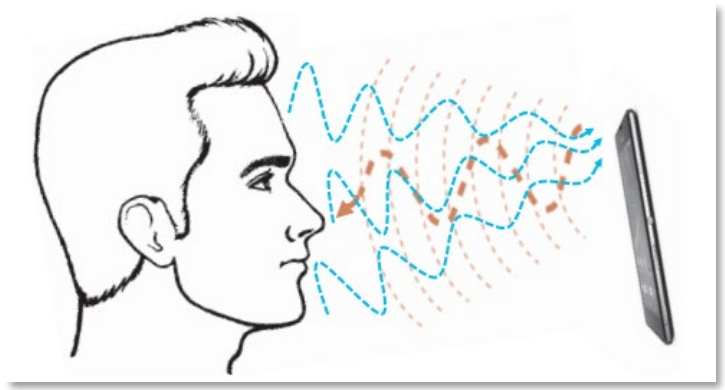
Mask blockage

Low/Unbalanced lighting

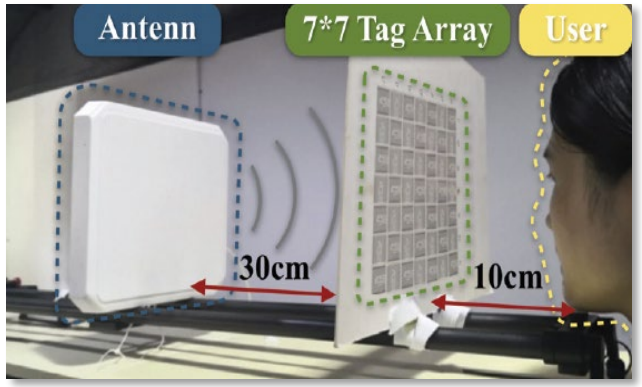
Privacy concern

Background

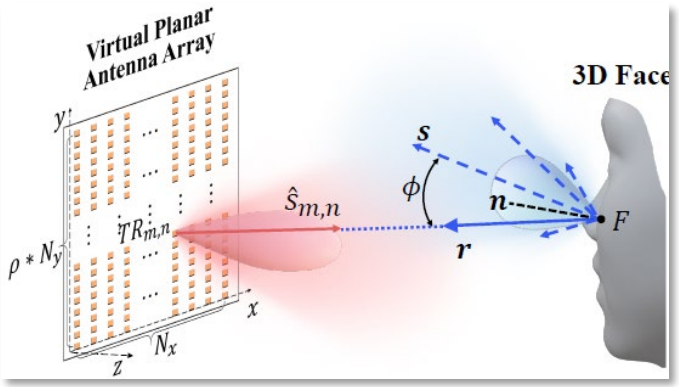
➢ **Facial recognition using facial reflected wireless signal**



EchoPrint – MobiCom 2018



RFace – INFOCOM 2021



mmFace – MobiCom 2022

- **Cannot resolve masked faces**
- **Requires visual assistance**
- **Relies on heavy hardware infrastructure**

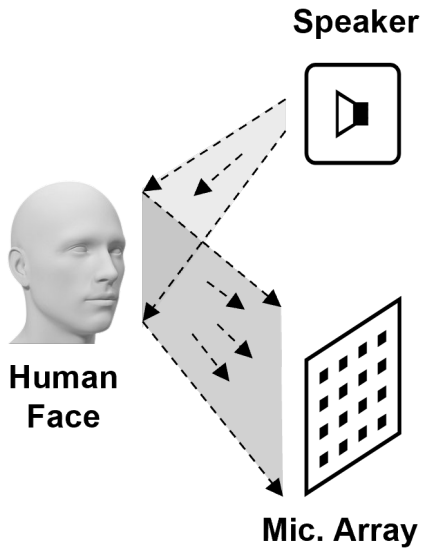
Motivation

- **Leveraging the special physical properties of acoustic signal for better sensing resolution and obstacle penetration capability.**

Acoustic signal's advantages of its physical properties (over RF signal)

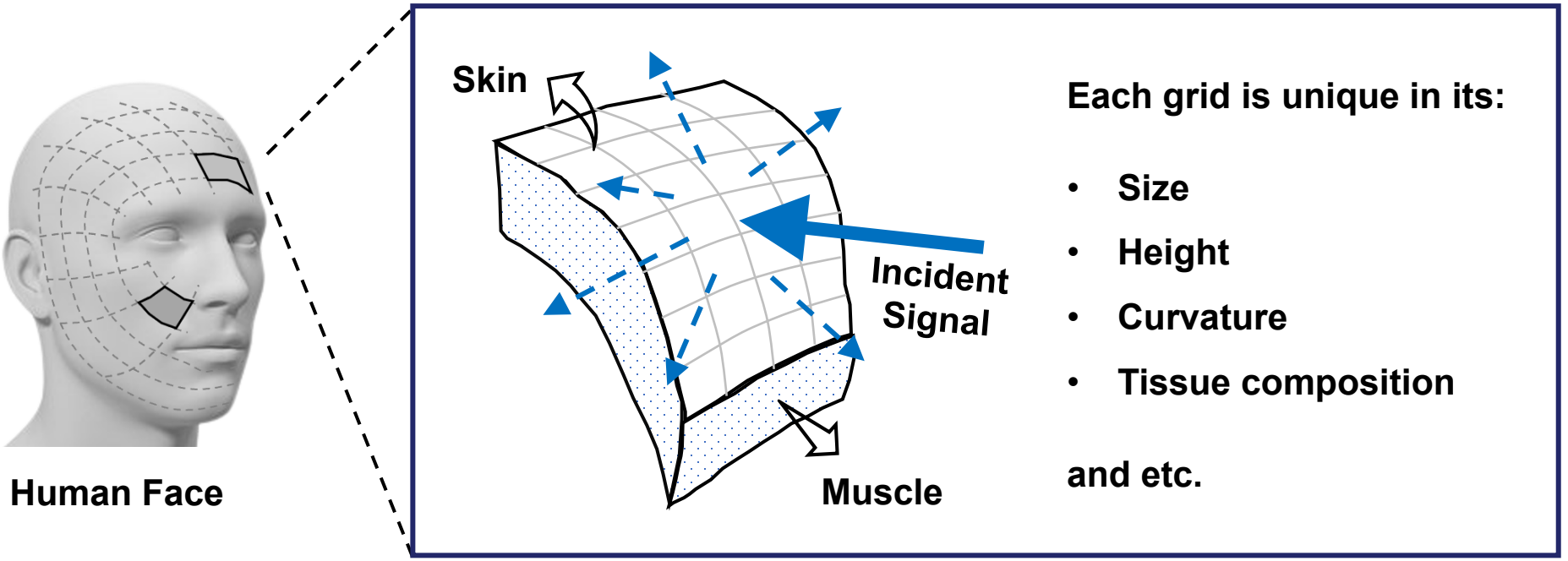
Lower propagation speed (340 m/s over the air)	⇨ Higher timing resolution
Lower carrier frequency (kHz – level)	⇨ Better obstacle penetration capability
Different propagation method (Mechanical wave)	

Acoustic Facial Recognition



Feature

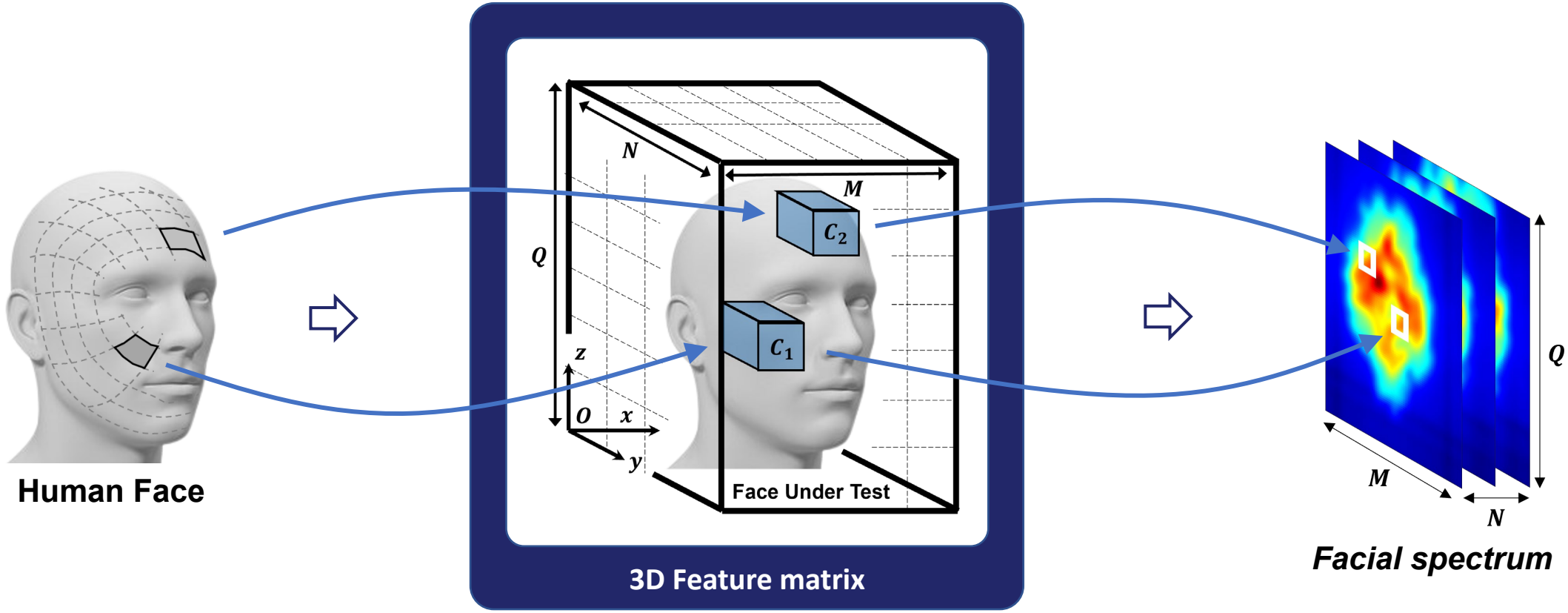
- **We exploit the spatial characteristics of human face for recognition.**



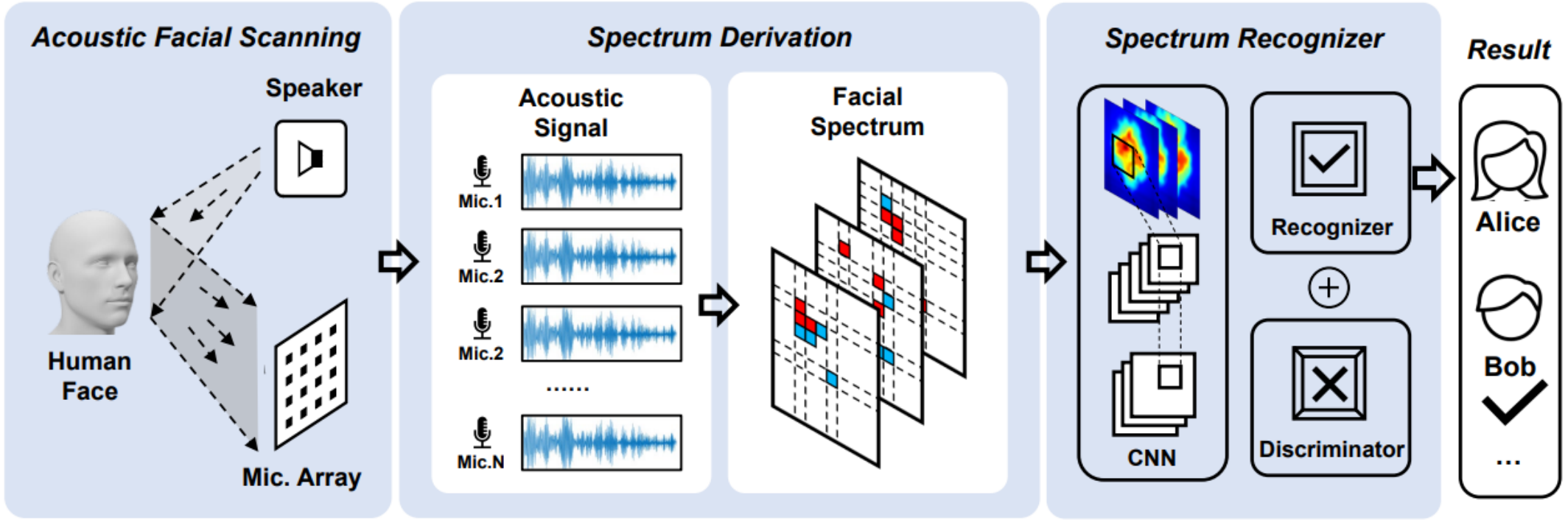
- **Reflection from all facial areas collectively forms a unique representation of the human face.**

Idea

- **Representing the spatial characteristic with *Facial spectrum*.**

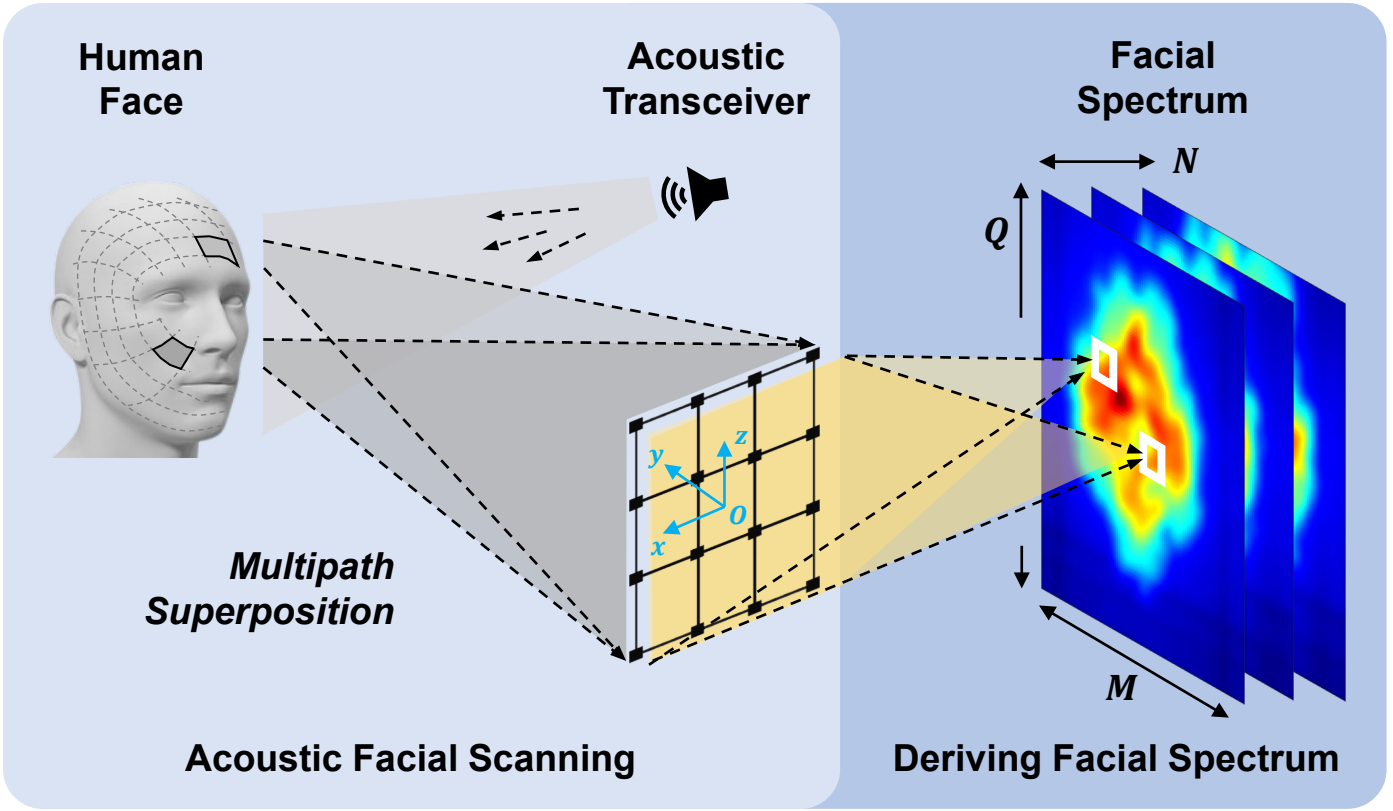


System overview



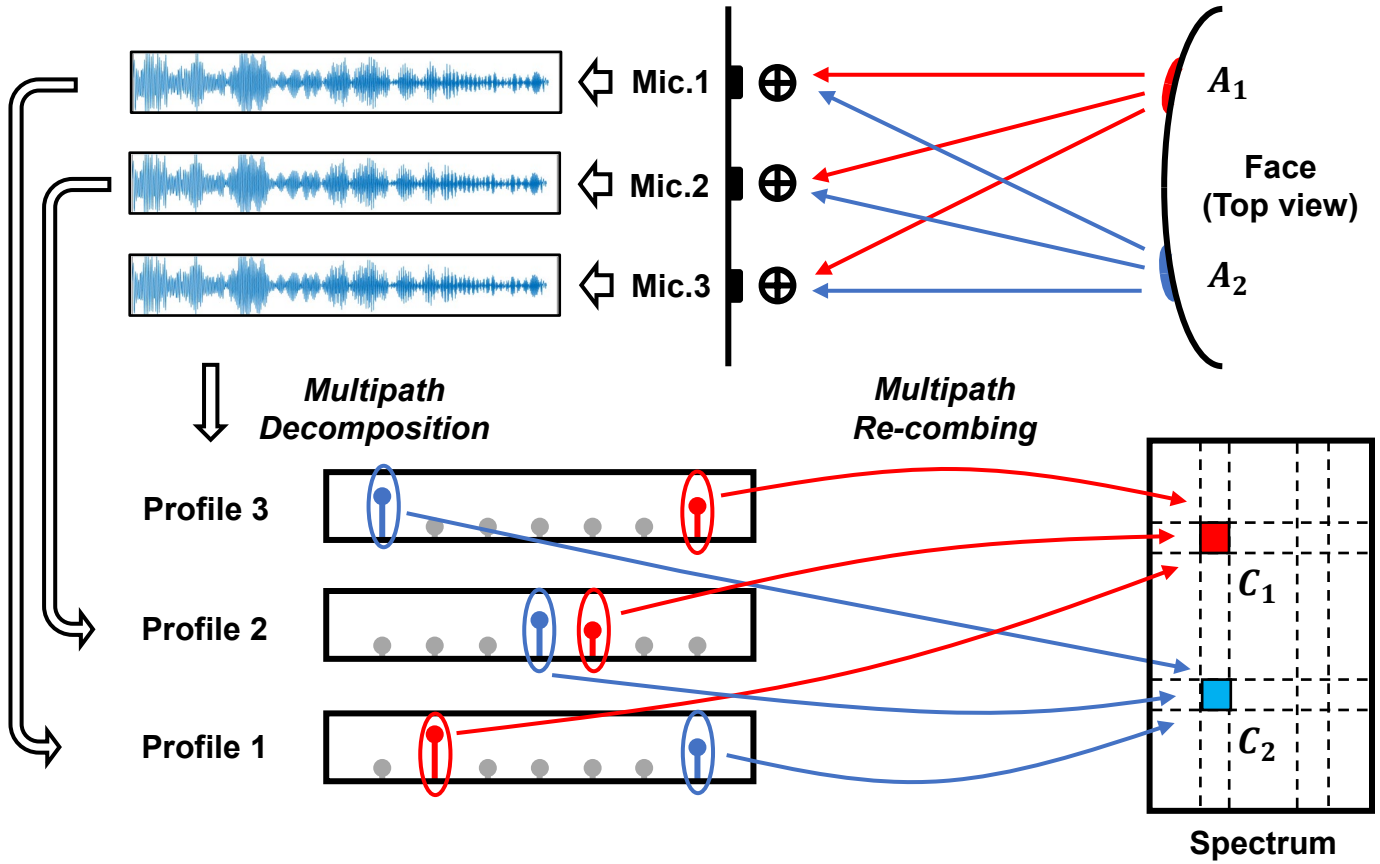
Challenge

- **1. How to resolve the reflected signals from different facial areas by using the received signals that are in a state of superposition?**



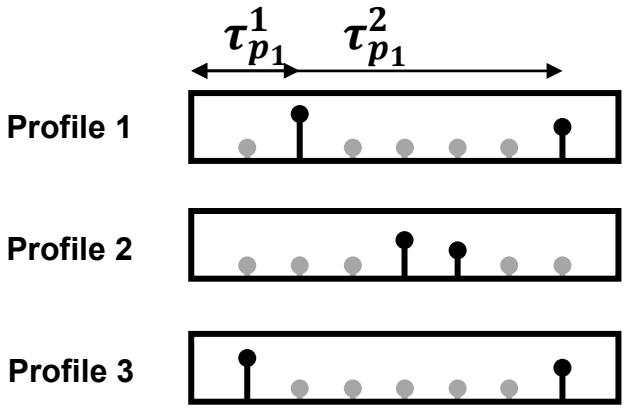
Design – Spectrum derivation

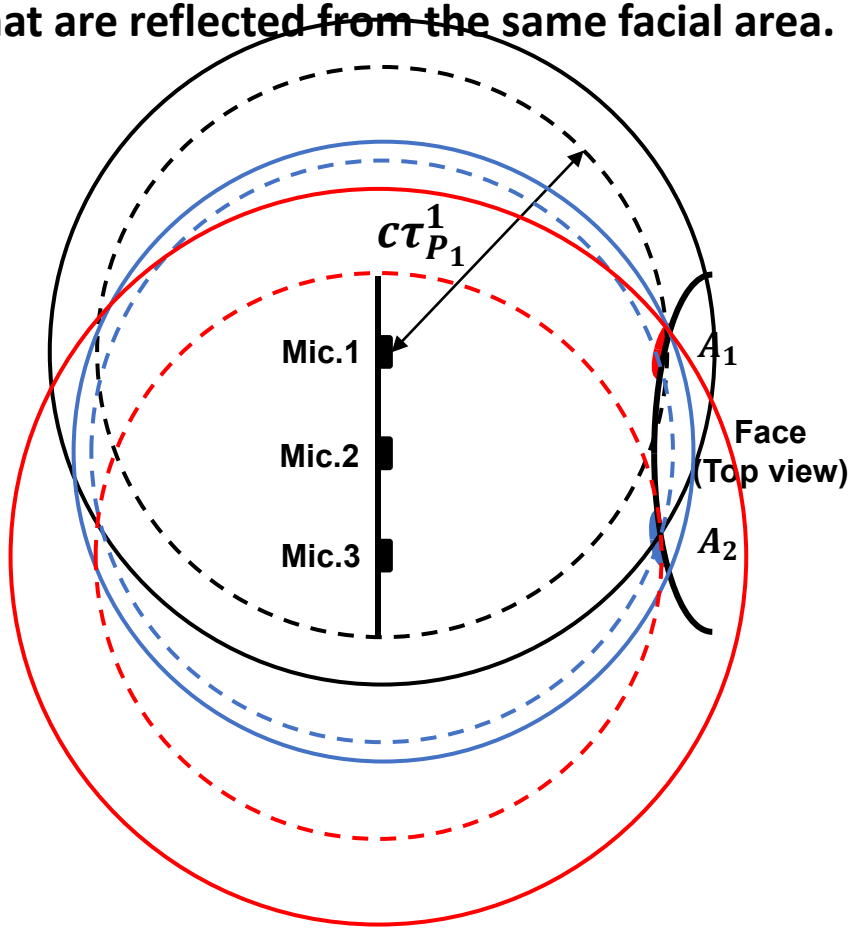
- **Basic idea: reversing the process from facial reflection to multipath superposition.**



Design – Spectrum derivation

- **Multipath Re-combining**
  - **Identifying the multipath components that are reflected from the same facial area.**





Design – Spectrum derivation

➢ Multipath Re-combining

- **Identifying the multipath components that are reflected from the same facial area.**

Criterion for multipath selection

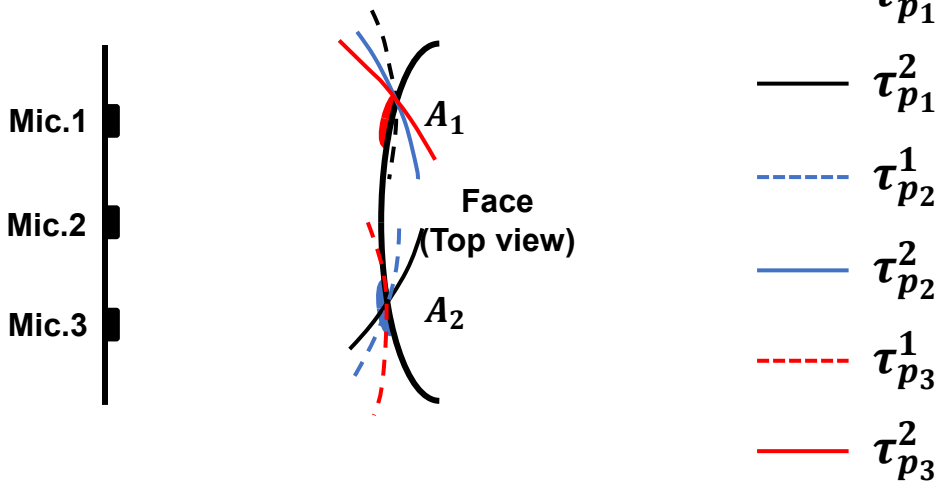
For the i-th facial reflecting area at the position $C_{F_i} = (x_{F_i}, y_{F_i}, z_{F_i})$, we select the k-th path from the profile of the j-th microphone

$$k = arg \min_{k \in Z^+} |c\tau_k - D_{F_i M_j} - D_{F_i S}|$$

where

$$D_{F_i M_j} = ||C_{F_i} - C_{M_j}||$$
$$D_{F_i S} = ||C_{F_i} - C_S||$$

C_{M_j} and C_S denote the position of the j-th mic. and the speaker, respectively



Design – Spectrum derivation

➢ Multipath Re-combining

Criterion for multipath selection

For the i-th facial scattering area at the position $C_{F_i} = (x_{F_i}, y_{F_i}, z_{F_i})$, we select the k-th path from the profile of the j-th microphone

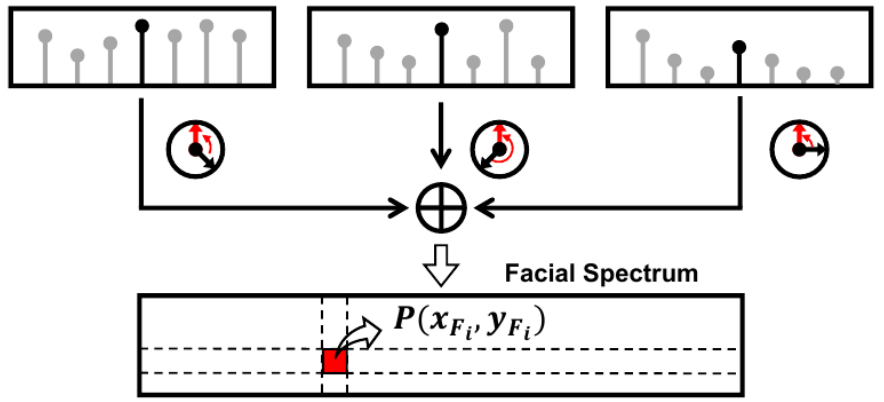
$$k = arg \min_{k \in Z^+} |c\tau_k - D_{F_iM_j} - D_{F_iS}|$$

where

$$D_{F_iM_j} = ||C_{F_i} - C_{M_j}||$$
$$D_{F_iS} = ||C_{F_i} - C_S||$$

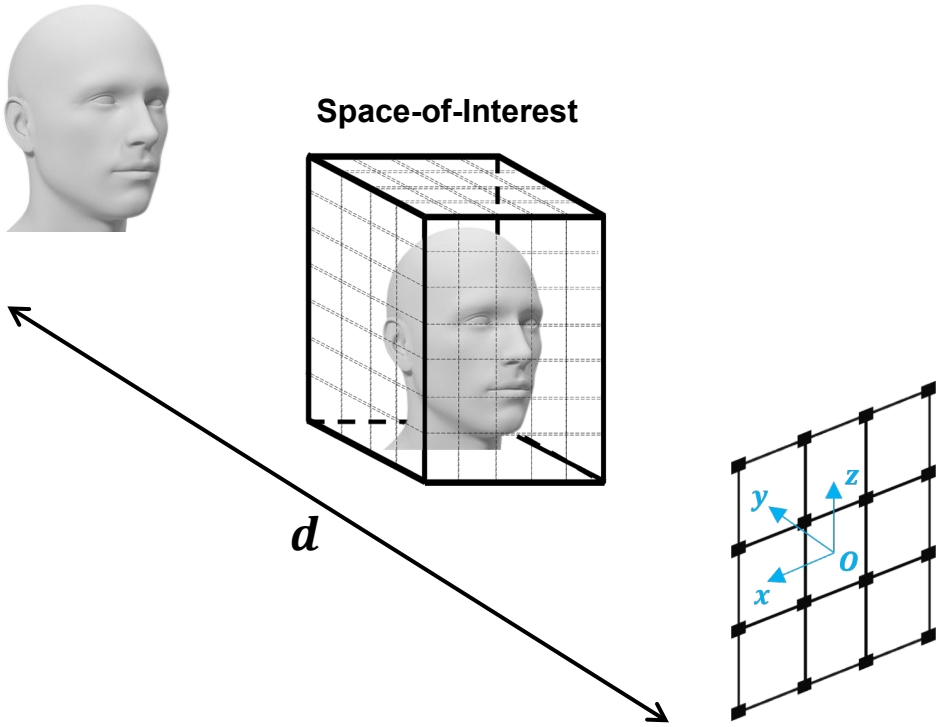
C_{M_j} and C_S denote the position of the j-th mic. and the speaker, respectively

Coherent combining



Design – Spectrum derivation

- **Locating the Space-of-Interest**

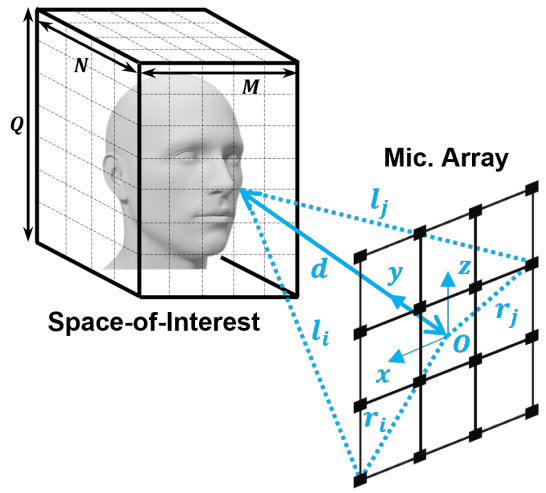


Design – Spectrum derivation

➢ Locating the Space-of-Interest

Estimating facial distance – a coarse-to-fine approach

Coarse estimation



$$d_{rough} = \sum_{i=1}^{M} d_i$$

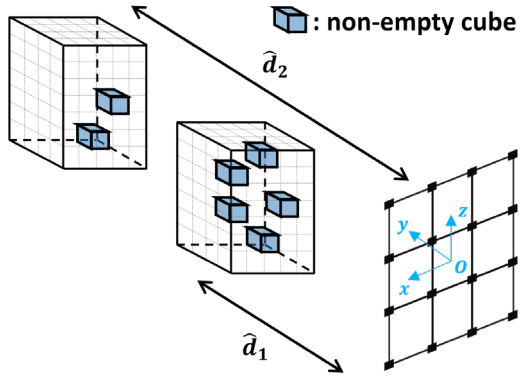
where

$$d_i = \sqrt{l_i^2 - r_i^2}$$

$$l_i \approx \frac{c\tau_{peak}}{2}$$

(an approximation)

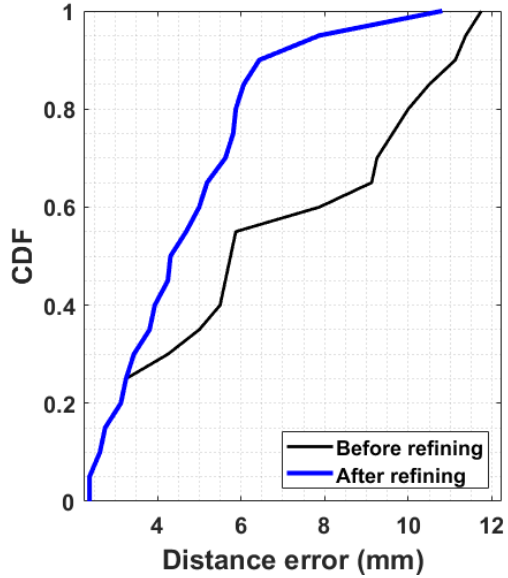
Refining the estimation



$$d_{fine} = \arg\max_{\hat{d}_i} N(\hat{d}_i)$$

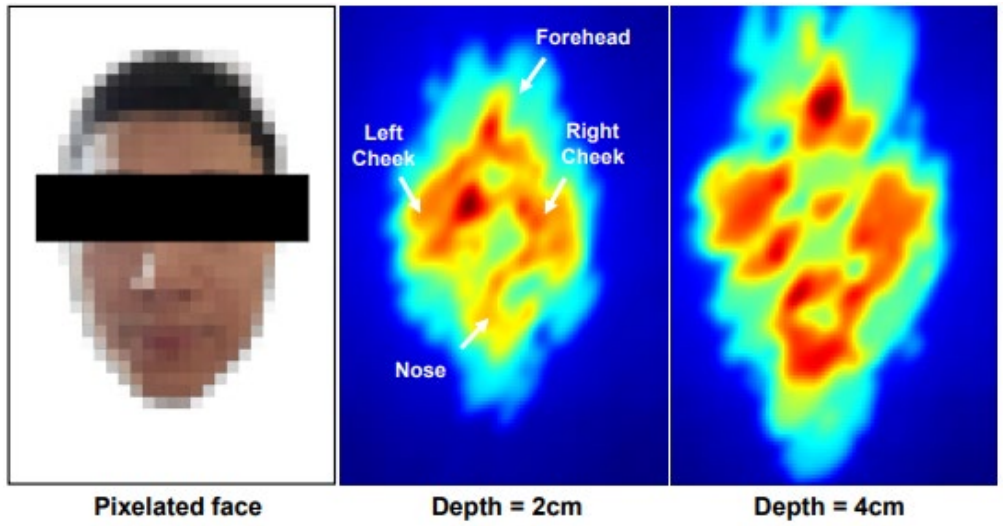
$$\hat{d}_i \in [d_{rough} - \frac{\eta}{2}, d_{rough} + \frac{\eta}{2}]$$

Benchmarking result

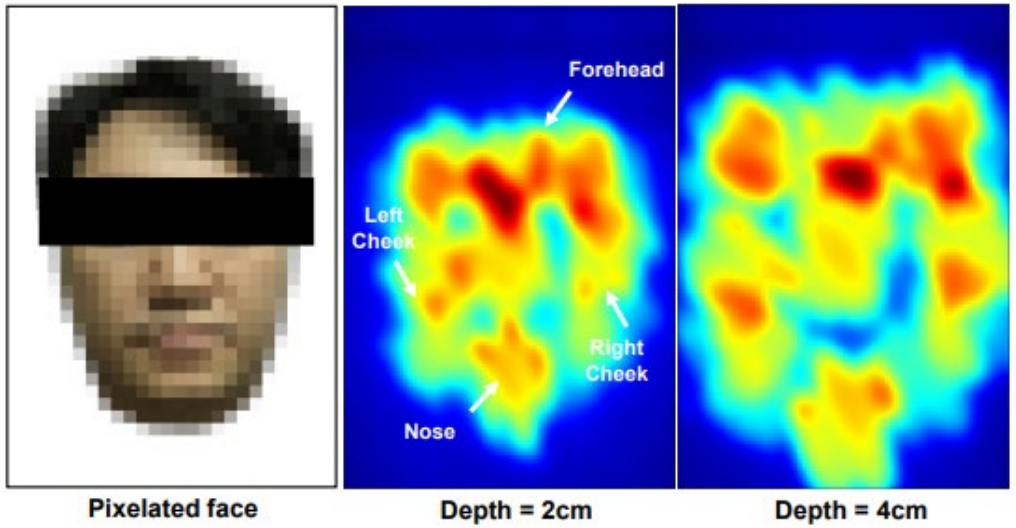


Design – Spectrum derivation

- **Showcase of the derived facial spectrum**



2D spectrum of User A (at different depth)

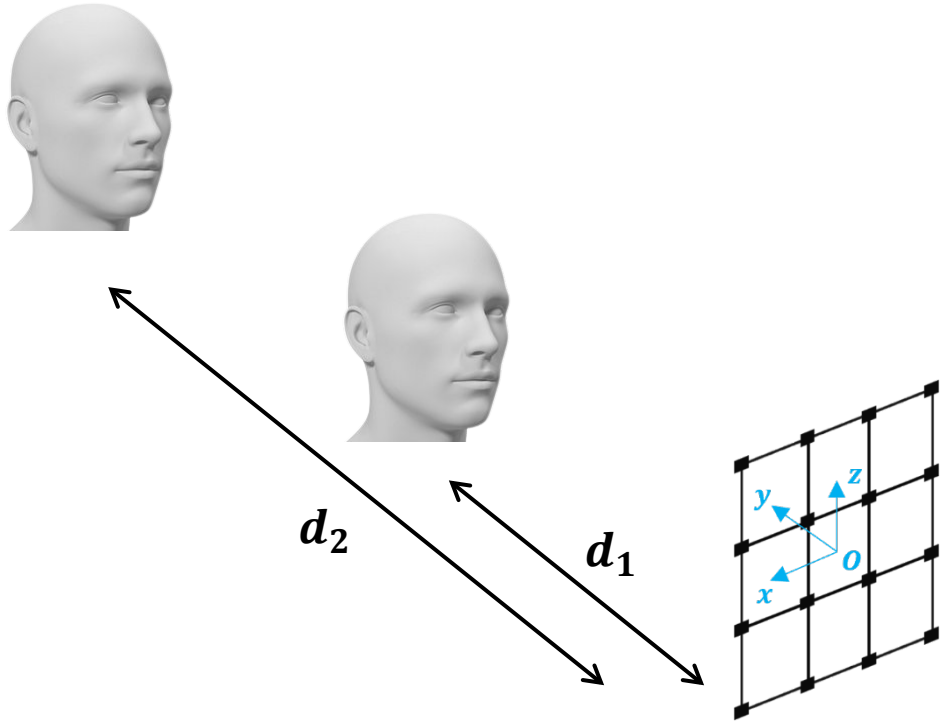


2D spectrum of User B (at different depth)

Challenge

- **2. How to avoid the impact of factors unrelated to identity?**

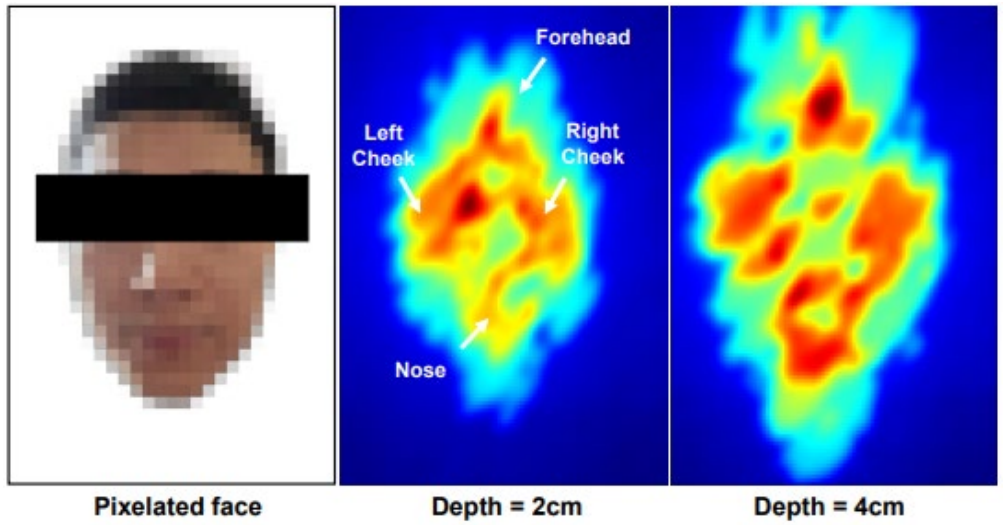
1) Facial mask blockage



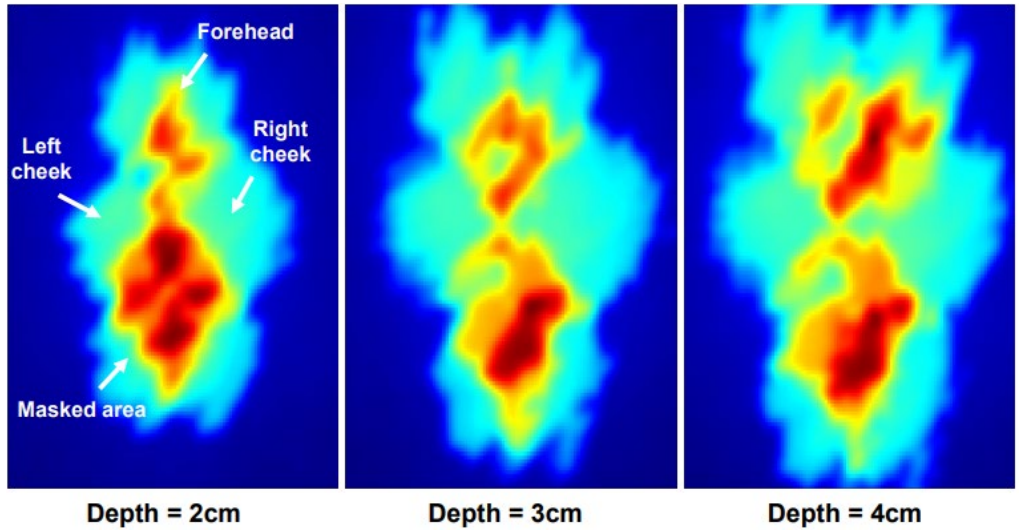
2) Facial – array distance

Design – Spectrum recognition

- **Facial mask varies the facial spectrum.**



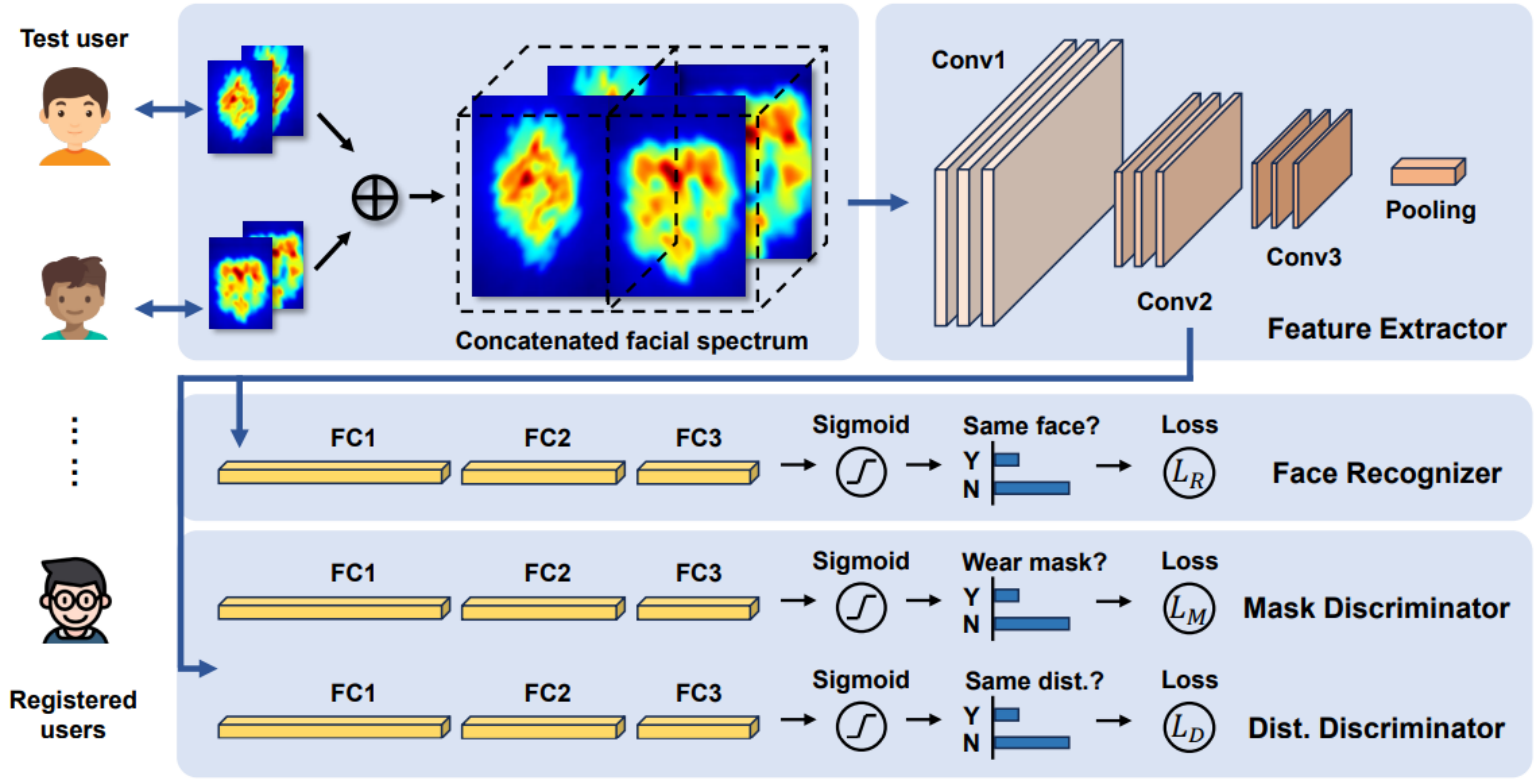
2D spectrum of User A, without mask



2D spectrum of User A, with mask

Design – Spectrum recognition

- **We design a RD-Net to provide accurate recognition even with facial mask blockage.**



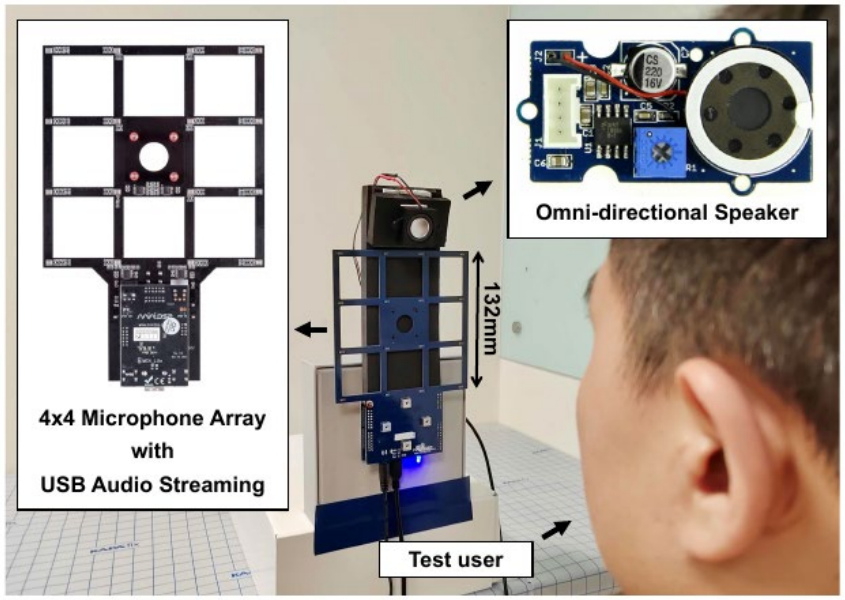
Loss function:

$$L = L_R - \frac{\alpha L_M + \beta L_D}{2}, 0 \leq \alpha, \beta \leq 1$$

Implementation

- **We implement AcFace with commercial low-cost acoustic hardware.**

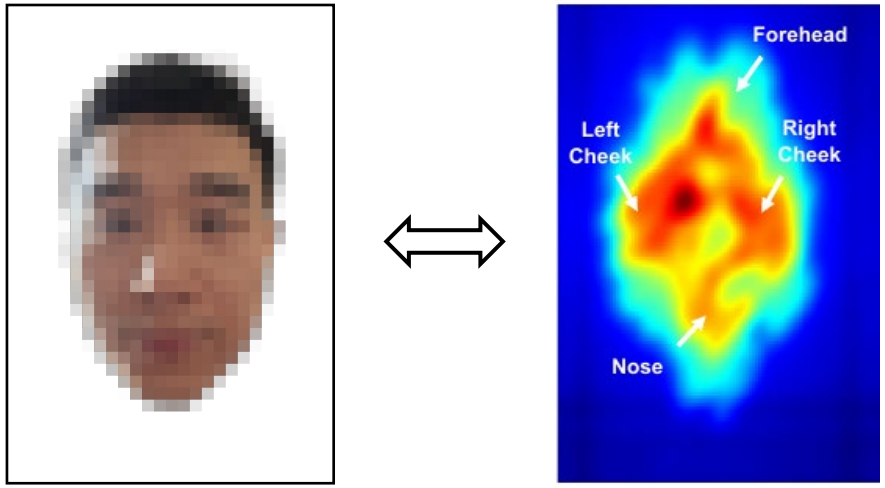
Prototyping the system



Hardware prototype and exp. scenario

Evaluation

- **Facial spectrum validation**



Facial image ⟺ **Facial spectrum**

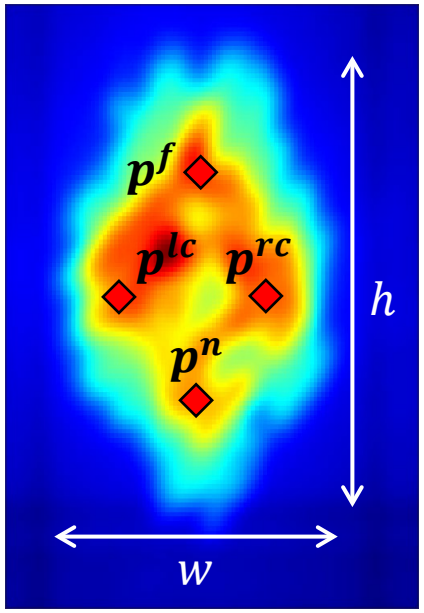
How effective does the spectrum represent essential facial features when compared with a facial image?

Evaluation

➢ **Facial spectrum validation**

- **How effective does the spectrum represent essential facial features when compared with a facial image?**

Defining *essential facial features*



- **Facial contour:** $\boldsymbol{s} = (w, h)$
- **Position of essential landmarks**
  - **Center of forehead:** $\boldsymbol{p^f} = (x^f, y^f)$
  - **Center of left cheek:** $\boldsymbol{p^{lc}} = (x^{lc}, y^{lc})$
  - **Center of right cheek:** $\boldsymbol{p^{rc}} = (x^{rc}, y^{rc})$
  - **Nose tip:** $\boldsymbol{p^n} = (x^n, y^n)$

$$\boldsymbol{F} = \{\boldsymbol{s}, \boldsymbol{p}\} = \{\boldsymbol{s}, [\boldsymbol{p^f}, \boldsymbol{p^{lc}}, \boldsymbol{p^{rc}}, \boldsymbol{p^n}]\}$$

Evaluation

➢ **Facial spectrum validation**

- **How effective does the spectrum represent essential facial features when compared with a facial image?**

Metrics

	Spectrum	Image
User k	$\boldsymbol{F}_{sp}^{k}$	$\boldsymbol{F}_{im}^{k}$
User j	$\boldsymbol{F}_{sp}^{j}$	$\boldsymbol{F}_{im}^{j}$

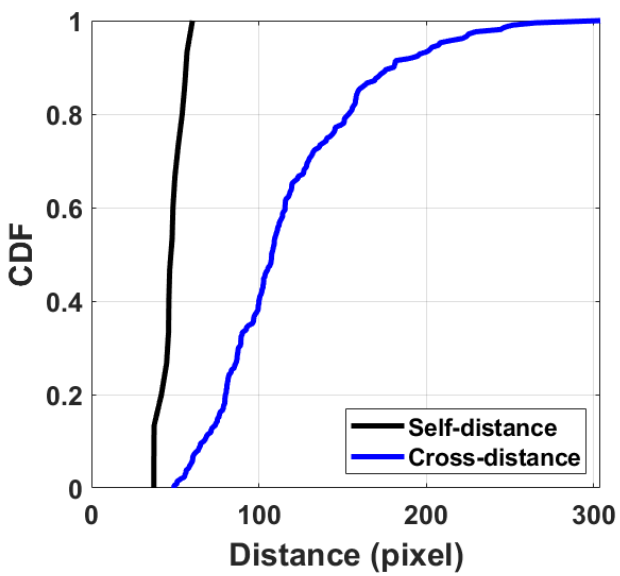
- **Self – distance:** $\boldsymbol{\delta}_k = |\boldsymbol{F}_{im}^{k} - \boldsymbol{F}_{sp}^{k}|$
- **Cross – distance:**

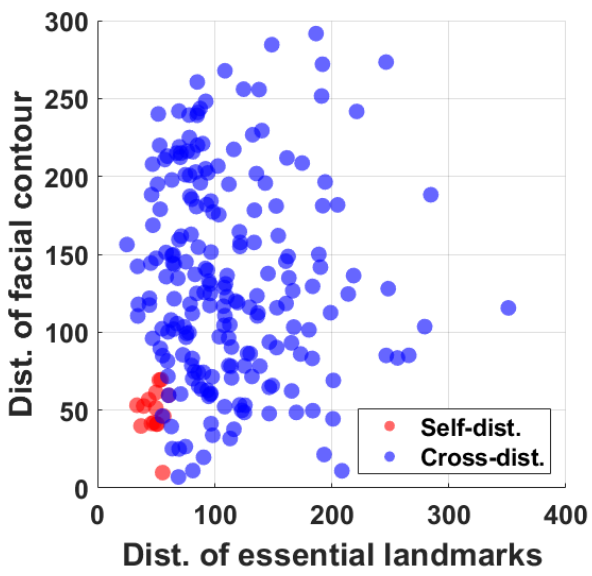
$$\boldsymbol{\Delta}_{kj} = \left|\boldsymbol{F}_{sp}^{k} - \boldsymbol{F}_{sp}^{j}\right| , \qquad k \neq j$$

Evaluation

- **Facial spectrum validation**
  - **How effective does the spectrum represent essential facial features when compared with a facial image?**

Results





Evaluation

- **End-to-end evaluation**

Comparative evaluation

Test setting	VGG-Face	FaceNet	SRT	AcFace
Without mask	98.05/97.73	**98.86 / 98.79**	98.81/97.06	95.88/96.12
With mask	83.16/83.25	85.63/86.66	**95.61 / 95.82**	**95.77 / 96.07**
With mask (dim)	77.67/77.32	78.11/79.67	81.57/83.79	**95.71 / 96.19**

Different environments

Environment	Precision (%)	Recall (%)	F1-score (%)	AA (%)
Meeting room	95.66/96.53	95.51/95.49	95.76/96.33	95.88/–
Lab	96.79/95.99	95.67/95.87	95.82/95.87	95.81/–
Office	95.29/95.90	94.82/95.83	94.66/95.86	95.45/–

Different number of users

Number of users	4	6	8	10	12	14	15
Ave. Accuracy (%)	98.81	98.67	96.72	95.88	95.97	96.13	95.67
Inference delay (ms)	16.62	21.54	26.92	31.36	34.80	39.96	43.39

Conclusion

- We propose acoustic facial spectrum, which can provide an accurate representing essential facial features of human faces.
- We devise a recognizer-discriminator network model to provide accurate and robust feature extraction/identification
- We prototype the system and conduct comprehensive real-world evaluation.

Check it out

Thanks