REsCUE: A framework for REal-time feedback on behavioral CUEs using multimodal anomaly detection

Riku Arakawa[†] and Hiromu Yakura[‡]

(equal contribution)

[†] The University of Tokyo, Japan
[‡] Teambox Inc. and University of Tsukuba, Japan
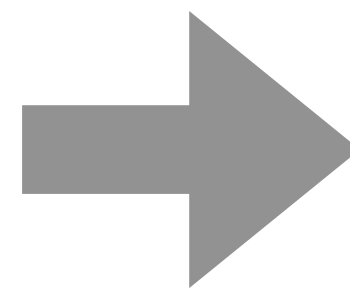
ACM CHI @ Glasgow, UK
May 7, 2019

teambox

THE
MENTALIST

Some people can infer human internal states through observation.

"**spontaneous & unregulated expression**"

Internal states
ex. impatient, nervous
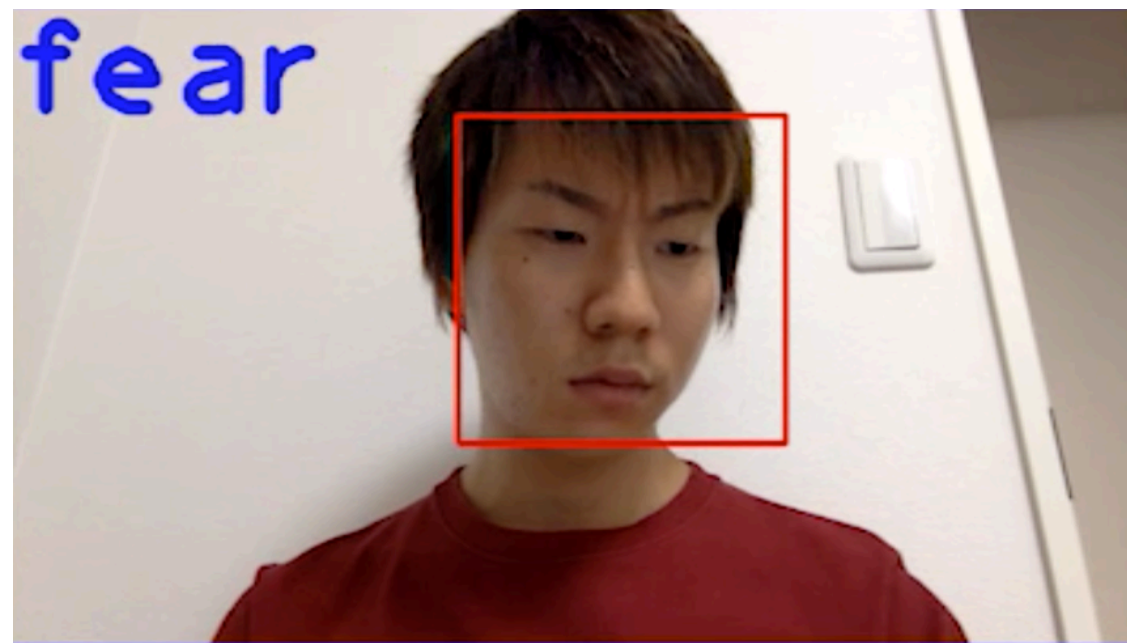
➡️

Nonverbal behavior
ex. arms crossed

[18] Charles Darwin, 1872. The Expression of the Emotions in Man and Animals. Jon Murray, London.
[19] Bella M DePaulo, 1992. Nonverbal behavior and self-presentation. Psychological bulletin.
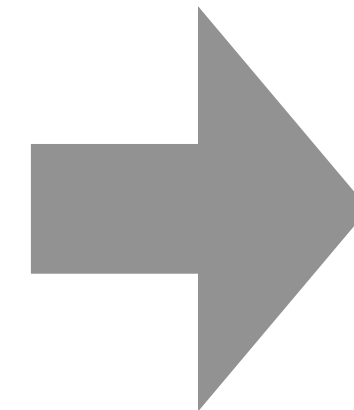
Development of human behavior analysis technologies

・ Facial expression recognition [14]
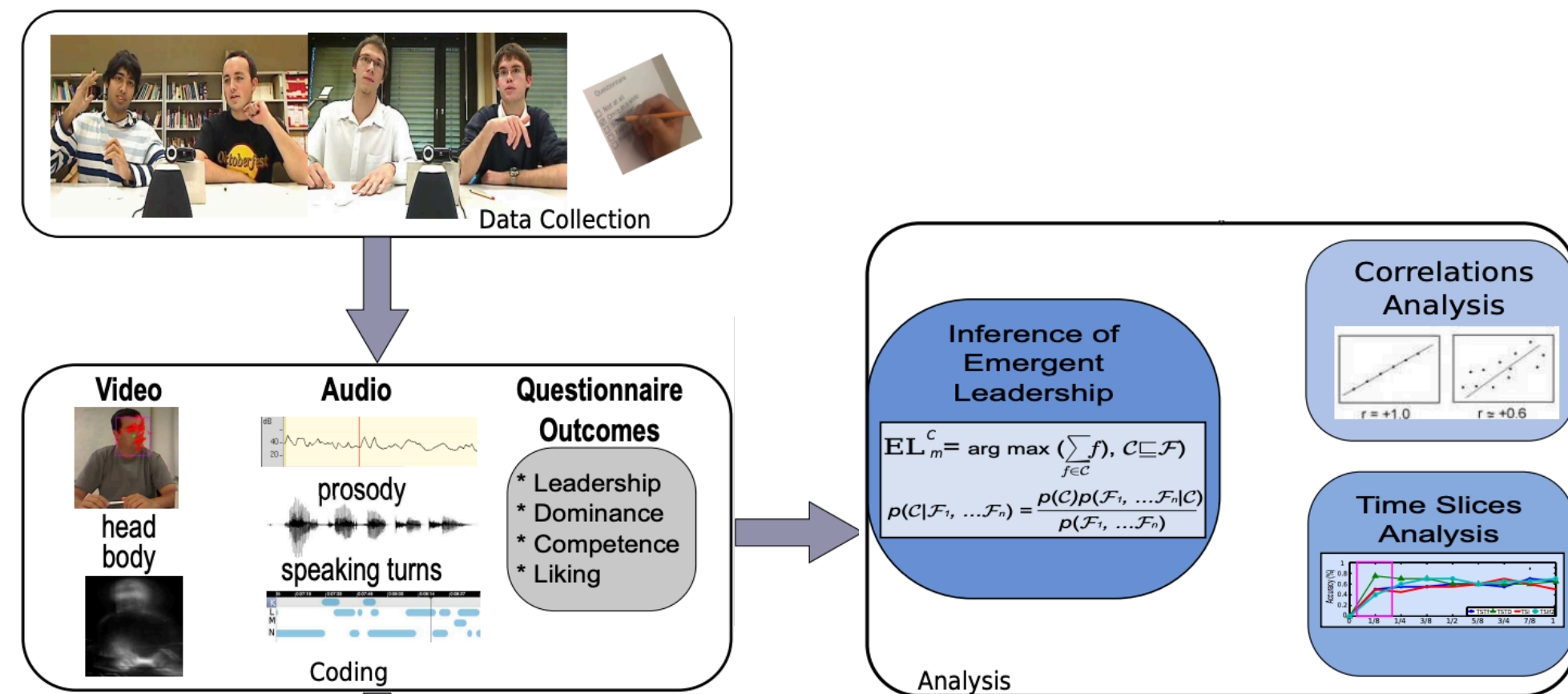

・ Posture estimation [61]

It is expected to analyze group or one-on-one conversation utilizing these techniques.

[14] Ilke Çugu et al., 2017. MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Frontal Face Images. arXiv.
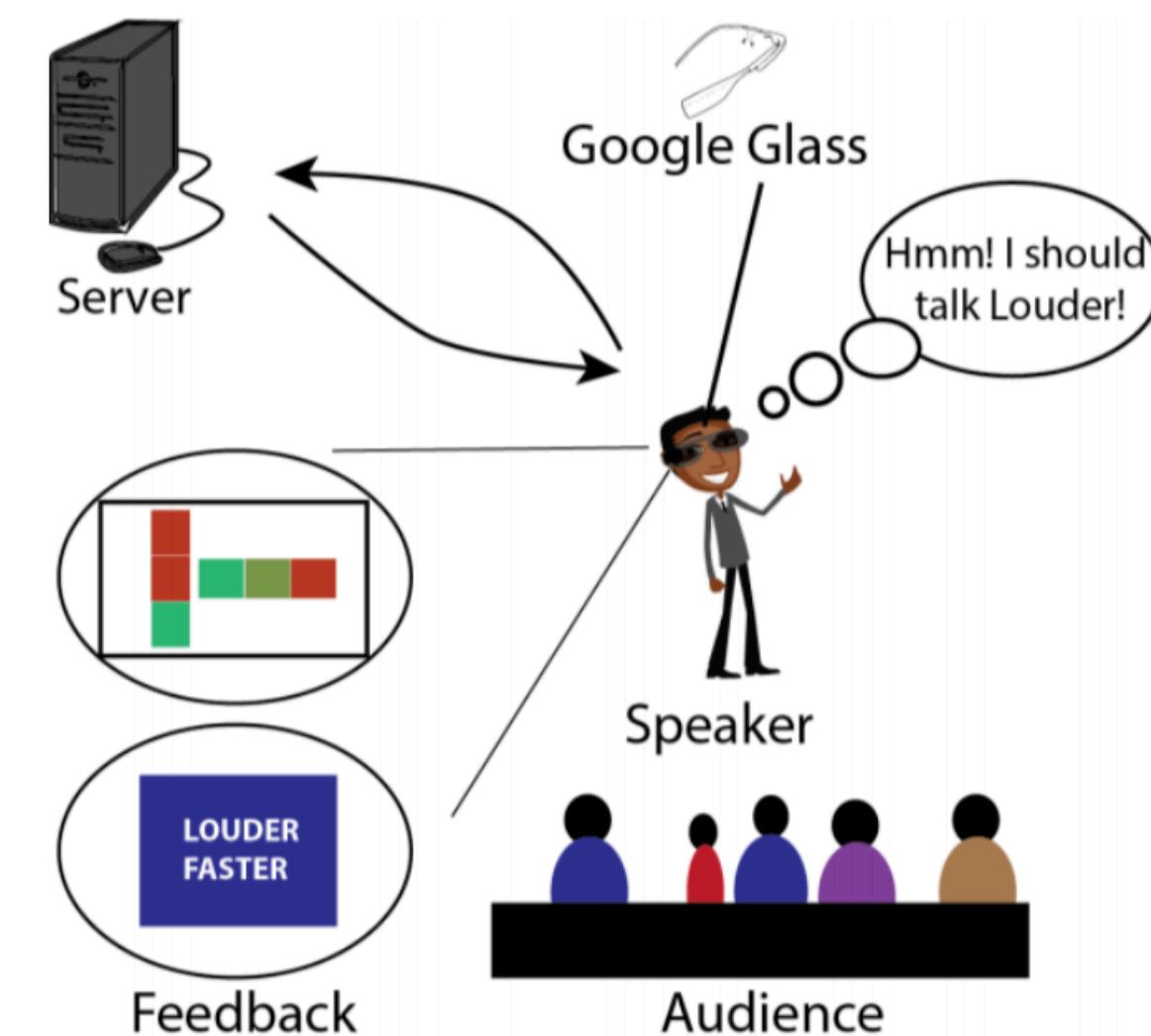[61] Shih-En Wei et al., 2016. Convolutional Pose Machines. IEEE CVPR.

Identify emergent leaders by analyzing human behaviors [51]

Apply rule-based method to pre-defined features like speaking length or head motion

Help public speaking [56]

Show feedback by words such as "louder" and "faster" via Google Glass comparing features with certain thresholds

Often assume heuristic rules so that computers can judge on each situations.

[51] Dairazalia Sanchez-Cortes et al., 2014. A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. ACM ICMI.
[56] Mohammad Iftekhar Tanveer et al., 2015. A Real-Time In-Situ Intelligent Interface to Help People With Public Speaking. ACM IUI.

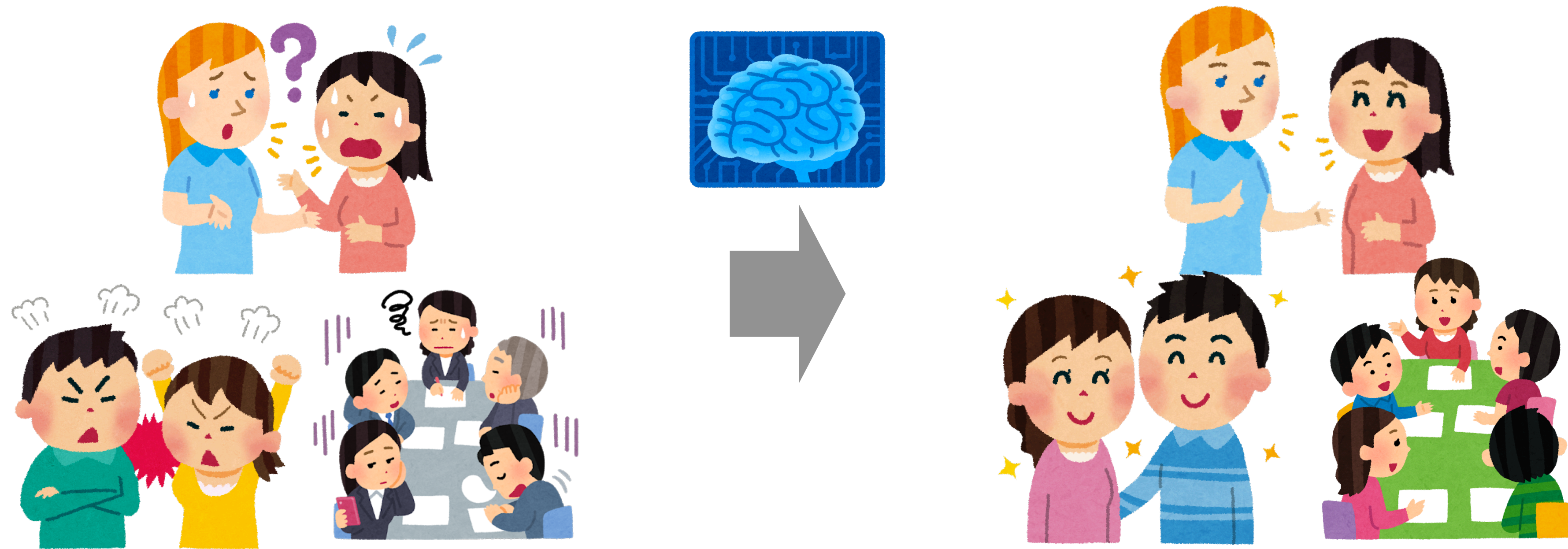Real conversations are not so simple, such as "active" or "loud".

It's nonsense to assume any rules for our conversation.

Can computers predict good cues for analyzing conversation *without any rules*?

**If computers can help us to analyze conversation that can be complex,
they can expand our communication in many situations.**

**As such a complex talking scene,
we chose *executive coaching* as our research object.**

Executive coaching

- is **one-on-one conversation** between a professional coach and a company manager to develop their professional skills and to achieve their goals.

- plays an important role in human resource development, which market share has increased to $2 billion [29].

[29] Robert G. Hamlin, et al. 2008. The emergent 'coaching industry': a wake-up call for HRD professionals. Human Resource Development International.

Research object: Executive coaching

Coaches

- ask smart questions to draw out coachees' opinions.

- are expected to notice a discrepancy between the verbal response and the actual thoughts using nonverbal cues [26].

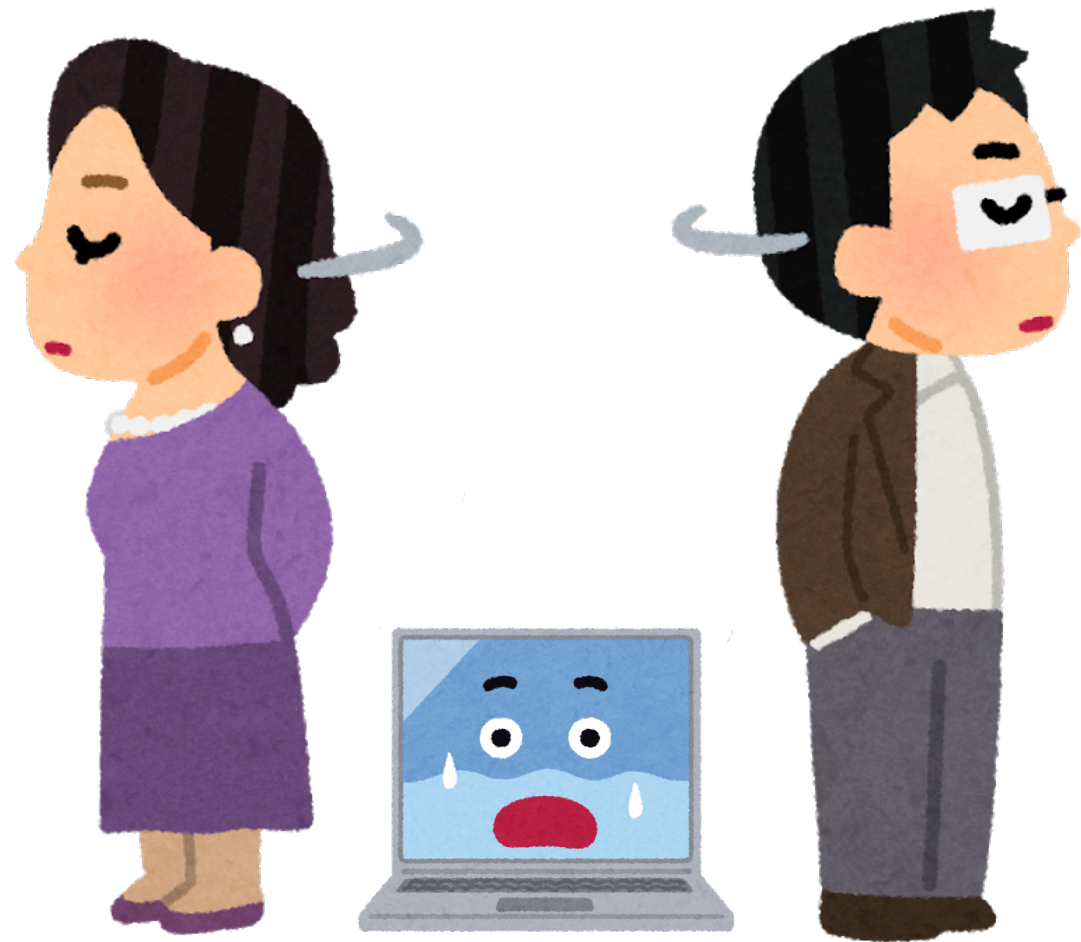**Coaches are required to be skillful not only at questioning but also at observation [6].**

**Coaching conversation is usually not so simple (smiling or not, for example).
Can conversational analysis help coaches during sessions?**

[6] Gary S. Bloom, et al. 2005. Blended Coaching: Skills and Strategies to Support Principal Development.
[26] Carol Kinsey Goman. 2008. The Nonverbal Advantage: Secrets and Science of Body Language at Work.

An insight: Observation and Judgement

Experienced coaches are not attracted by current conversation analysis technologies.

NG: Computers only can observe and judge on states.

Human
Pros: Good at understanding context
Cons: Difficult to keep stable perspective
due to subjectivity or mental load

Computers
Pros: Stable performance
Cons: Not good at deal with context

**Once an output from computers contradicts their intuition,
coaches completely ignore computers.**

**Separation of observation and
judgment would be an
alternative way of cooperation.**

Proposed method: REsCUE

1. **Overview: Real-time feedback using anomaly detection**

2. Anomaly detection algorithm

3. Visualization based on GMM

4. Feedback design

**To realize the separation of observation and judgement,
our solution is unsupervised anomaly detection and human interpretation.**

The system detects nonverbal cues of the coachee and notify to the coach in real-time.

Detailed workflow

1. Extract posture and gaze information of the coachee.

2. Calculate outlierness score using anomaly detection algorithm.

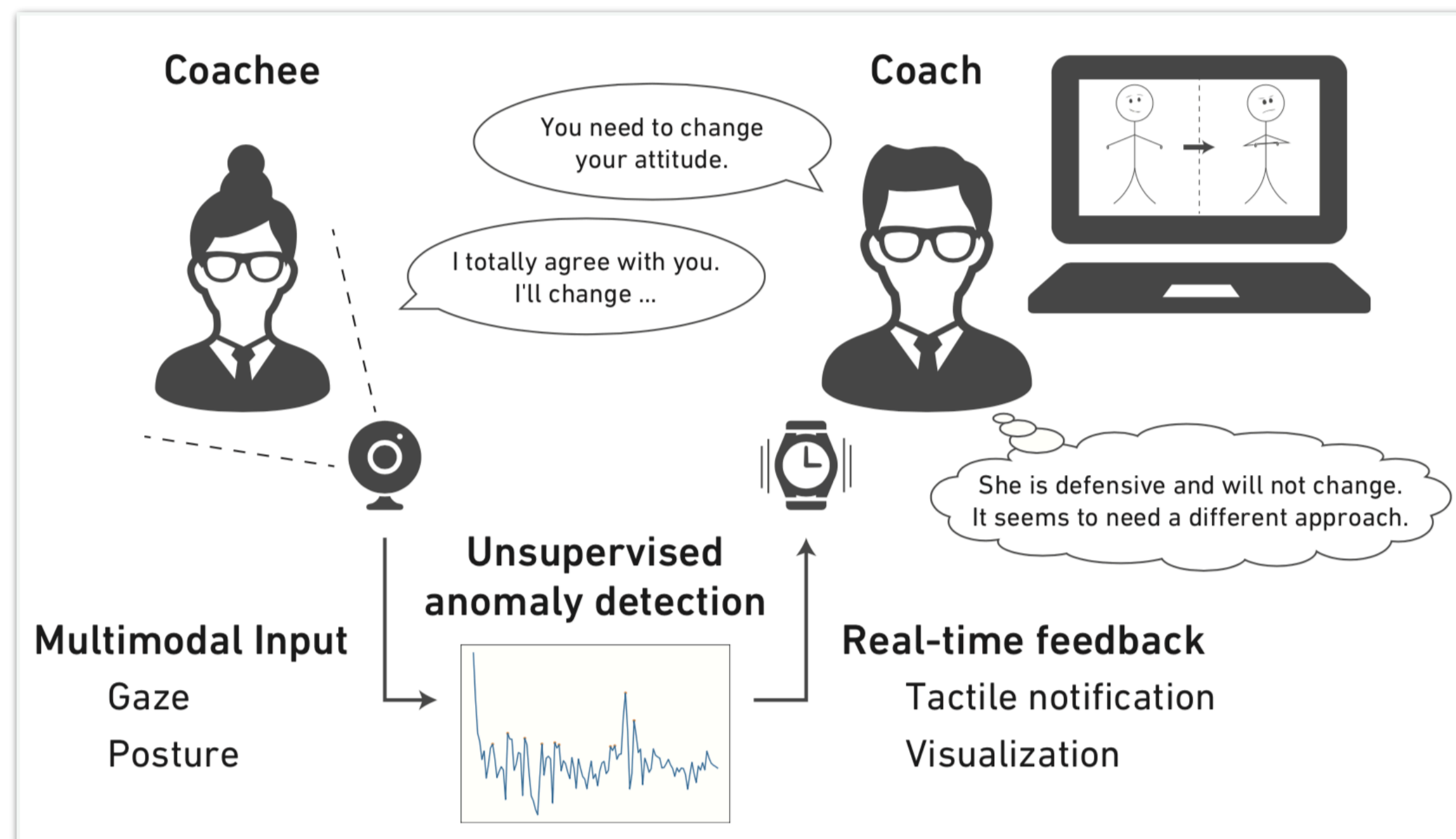3. Notify the coach in real-time with an intepretive visualization.

Proposed method: REsCUE

1. Overview: Real-time feedback using anomaly detection

2. Anomaly detection algorithm

3. Visualization based on GMM

4. Feedback design

2. Anomaly detection algorithm

Our system expects these points for the detection algorithm:

- **Unsupervised** detection: to rule out arbitrary criteria or assumptions
- **Real-time** detection: to give feedback to the coach during the session
- **Interpretive** detection: to help the coach understand the internal states

We use an algorithm based on **time-adaptive gaussian mixture model** [61].

- Our GMM gradually adapts to newly obtained nonverbal behavior data.
  - Then, **when the trend of the input data suddenly changes, we can detect it from the lowering of its likelihood.**

[61] Kenji Yamanishi, et al. 2004. On-Line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms. Data Mining and Knowledge Discovery.
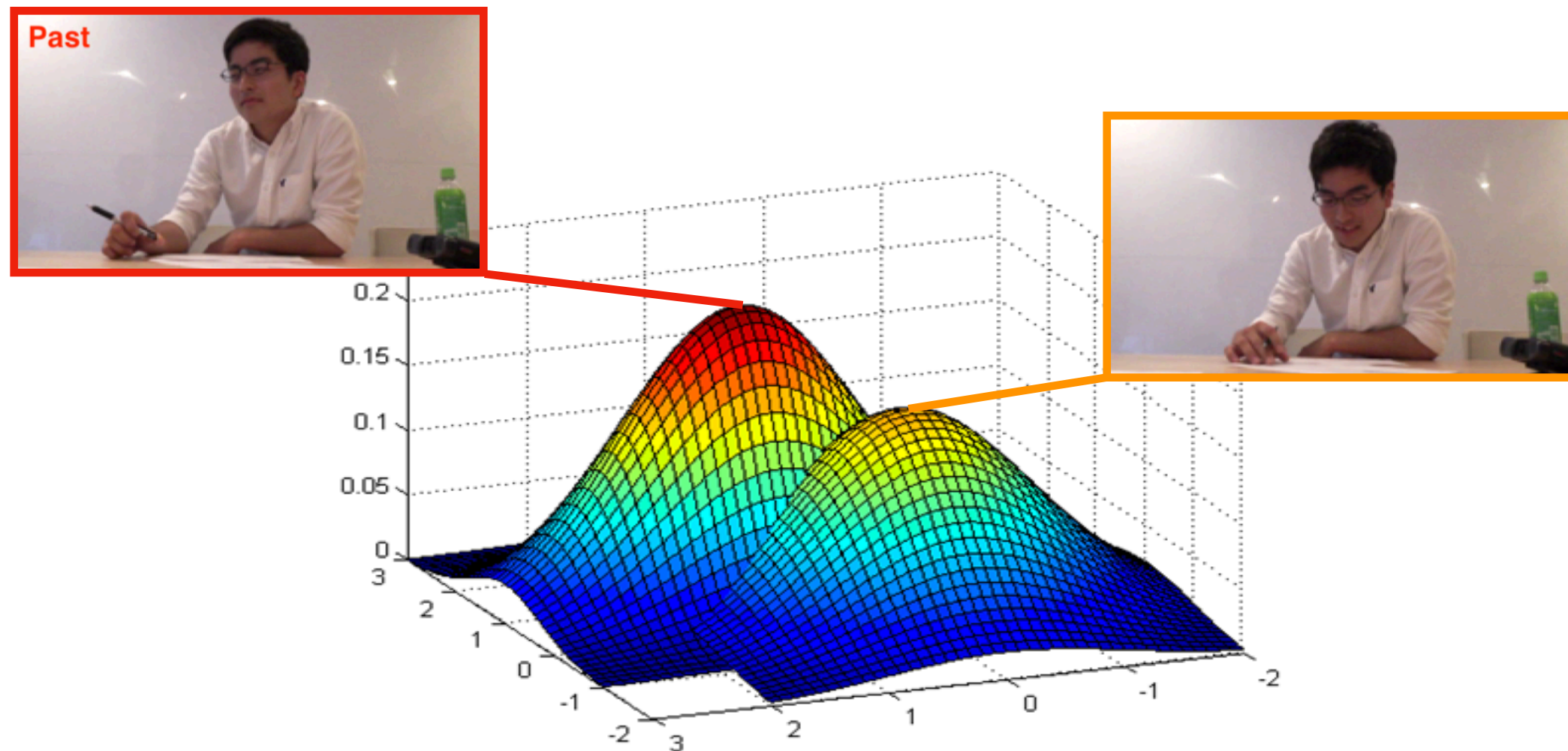
Proposed method: REsCUE

1. Overview: Real-time feedback using anomaly detection

2. Anomaly detection algorithm

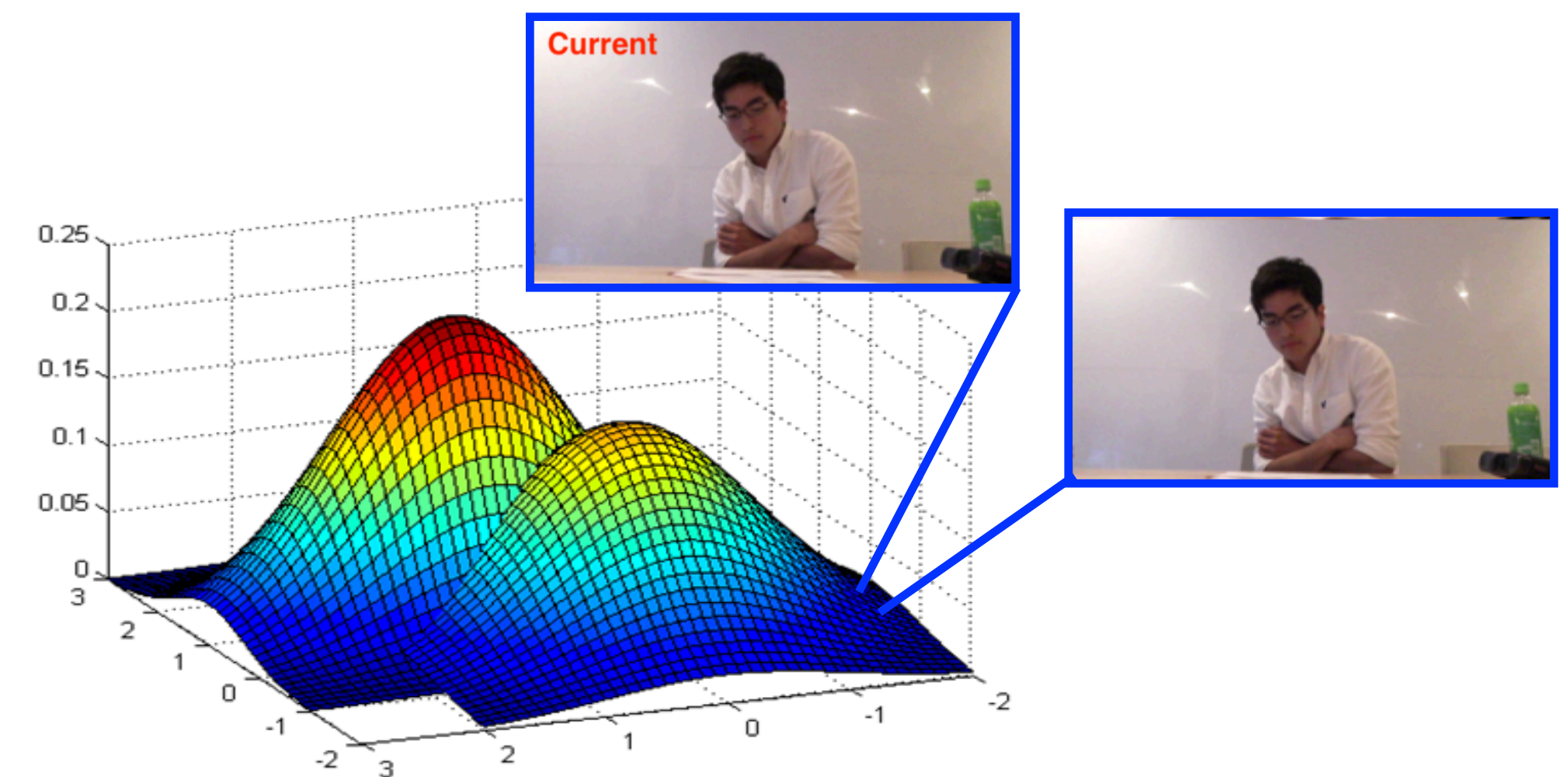3. **Visualization based on GMM**

4. Feedback design

3. Visualization based on GMM

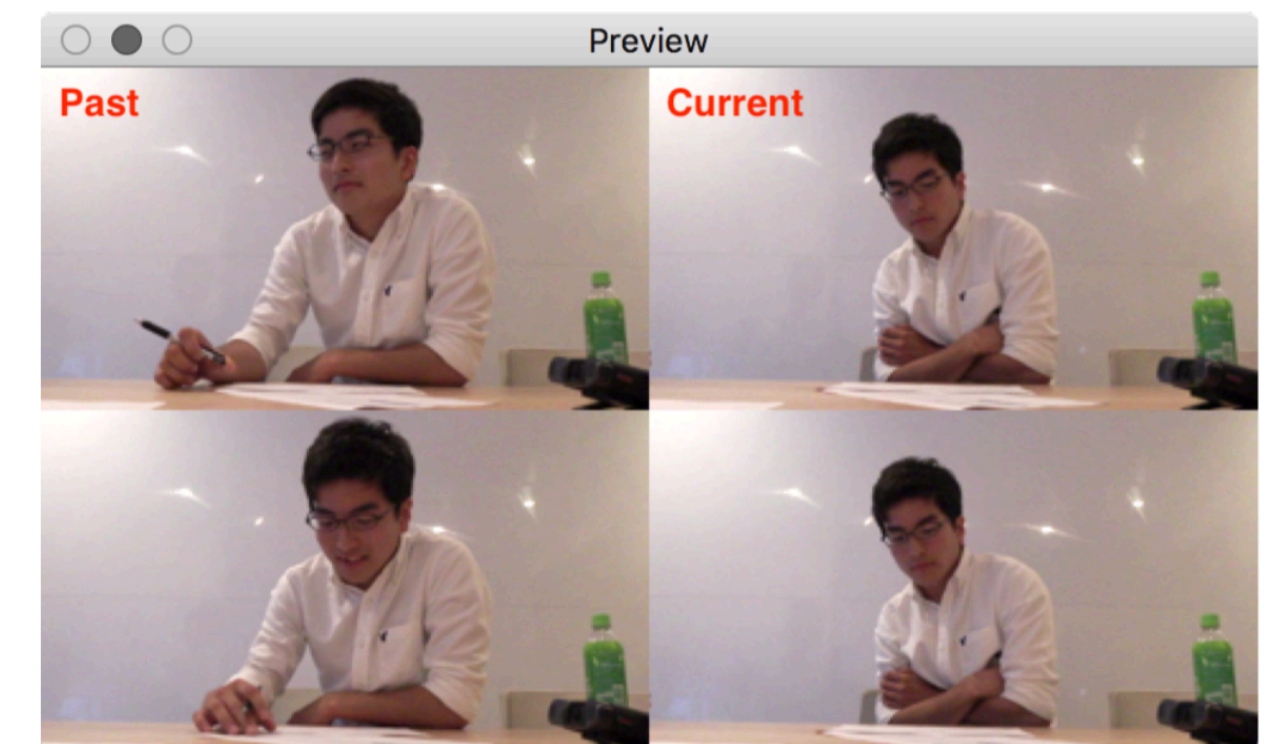We also exploits GMM to provide interpretative visualization.

In GMM, **each component shows the past representative states.**

When we have a new data with low likelihood, it would be anomaly.

Just by arranging these frames, the coach can compare them and understand the change easily even during the session.

Proposed method: REsCUE

4. Feedback design

Considering that the coach uses our system during the conversation,
the real-time feedback should not be interruptive.

We exploit **tactile feedback** via **smart watch**.

↳ Tactile feedback is known not to interfere concurrent task [20]
even during social interactions [15].

Our system notifies when the outlierness exceeds a certain threshold
while presenting the outlier frames on their PC.

To fully leverage the interactiveness of smart watch,
we enable the coach to adjust the threshold during the session.

[20] Aaron E. Sklar, et al. 2000. Good Vibrations: (cont). Human Factors.
[15] Ionut Damian, et al.. 2016. Exploring the Potential of Realtime Haptic Feedback during Social Interactions. TEI.

We show a simplified example
to illustrate **REsCUE**'s assist for coaches.

Preliminary experiment: Overview

Purpose
- Verify the availability of the proposed method.
- Determine which input modality is suitable to detect cues.

Procedure
1. Participated coach had a session while recording video.
2. The coach watched the video and list the top 10 most important cues of the coachee to infer internal states.
3. We applied the proposed method for each combination of input modalities from posture, gaze, facial expression.
4. We compared the annotated cues with the detected cues based on metrics such as recall and Kendall's τ-ranking distance.
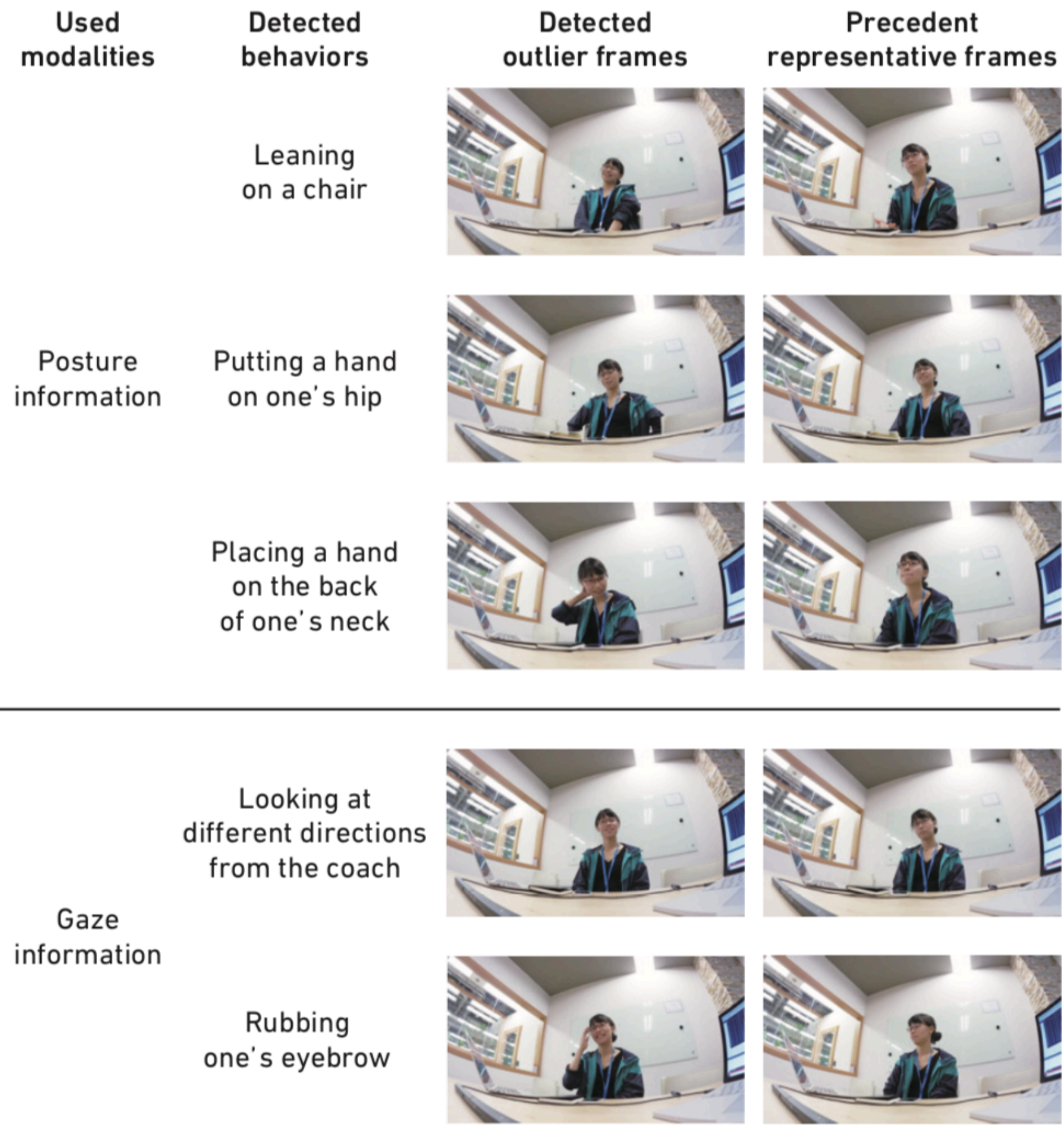
Data   6 sessions with 3 professional coaches (4h 28m 35s in total)

Combination of the posture and gaze information showed the best performance.

Used modalities			Metrics	
Posture	Gaze	Facial expression	Recall	Average of τ distance
✓			0.57 ± 0.08	0.42 ± 0.14
	✓		0.38 ± 0.15	0.21 ± 0.12
		✓	0.08 ± 0.10	0.01 ± 0.02
✓	✓		$\mathbf{0.68} \pm 0.08$	$\mathbf{0.64} \pm 0.09$
✓		✓	0.57 ± 0.12	0.40 ± 0.16
	✓	✓	0.37 ± 0.12	0.20 ± 0.12
✓	✓	✓	0.67 ± 0.16	0.61 ± 0.12

In addition, the coach regarded some detected cues as informative even though it is not listed when they watched the video.

← From the posture and gaze, we obtained these various behavioral cues, for example.

[14] Ilke Çugu, et al. 2017. MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Frontal Face Images. arXiv.
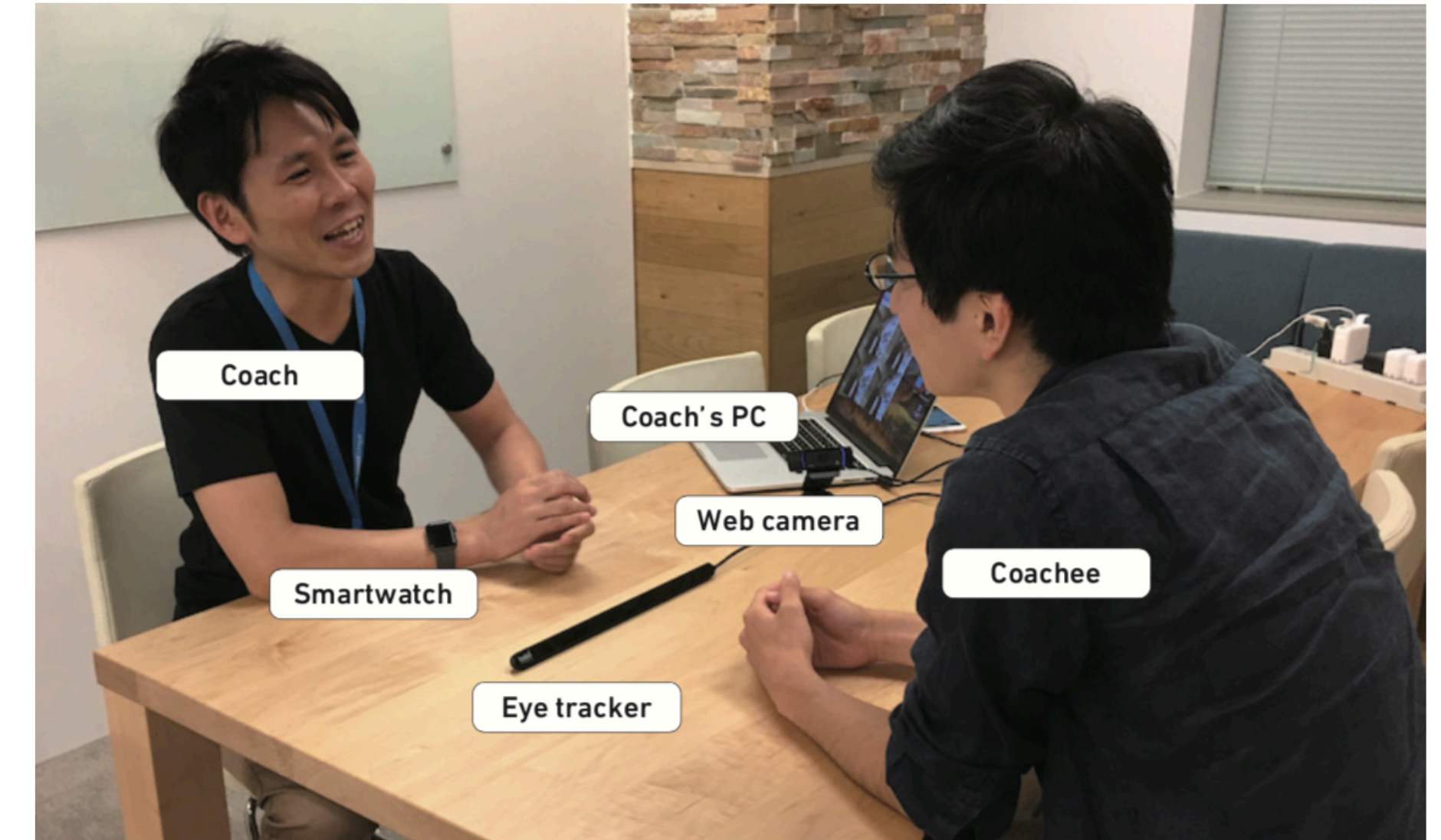
User study: Overview

Purpose

- To verify the effectiveness of the proposed system including the real-time feedback

Procedure

1. Participated coach had a session using the system.

2. We conducted short interviews with the coach in these points:

   - Usability

   - Effect of the feedback regarding the quality of the session

Coach
Coach's PC
Web camera
Smartwatch
Coachee
Eye tracker

Data 15 sessions with 5 professional coaches

We found **all of the participating coaches responded positively** in the interview.

User study: Usability

One participated coach commented its advantage on feedback design:

- Simpler feedback such as just showing "defensive" could also be easy to understand.
- However, **if it contradicts my feelings, I could get confused and might ignore the feedback**.
- In that respect, **this system passes the initiative to me and does not cause such confusion**.

Another coach gave us suggestions for future improvements:

- Further inferences are possible if this detects similar past behaviors or shows the trend in the behavioral cues.

User study: Effect of the feedback on the session

Many coaches commented that they changed the course based on the feedback:

I had been convinced that the coachee was agreeing to my proposal, **but from the given feedback, I noticed that it didn't seem true**.

So, I was able to make a decision to explain my proposal more carefully until he was satisfied.

I was impressed when the smartwatch vibrated immediately after I asked a delving question having butterflies in my stomach.

From the feedback, I became convinced that the underlying cause of the current issue lies there, and succeeded in having a deep discussion in a short period of time.

Conclusion & Future direction

- We introduced **REsCUE**, an intelligent system for use in coaching sessions.

    - It can automatically **detect nonverbal behavioral cues** of coachees.

    - It then provides feedback to coaches in real-time.

- From comments in actual coaching sessions,
  it is suggested that **REsCUE help coaches converse with coachees**.

We remark that **REsCUE does not require any prior knowledge or rules**
and can be used outside coaching sessions.

We expect that such unsupervised approaches go mainstream
to deal with the complexity of human in social interactions.

Riku Arakawa
The University of Tokyo
https://rikky0611.github.io

Hiromu Yakura
Teambox / University of Tsukuba
https://yumetaro.info

slide: https://bit.ly/REsCUE