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

**Riku Arakawa**[*]
University of Tokyo
Tokyo, Japan
arakawa-riku428@g.ecc.u-tokyo.ac.jp

**Hiromu Yakura**[*†]
Teambox Inc.
Tokyo, Japan
hiromu@teambox.co.jp

## ABSTRACT

Executive coaching has been drawing more and more attention for developing corporate managers. While conversing with managers, coach practitioners are also required to understand internal states of coachees through objective observations. In this paper, we present *REsCUE*, an automated system to aid coach practitioners in detecting unconscious behaviors of their clients. Using an unsupervised anomaly detection algorithm applied to multimodal behavior data such as the subject's posture and gaze, REsCUE notifies behavioral cues for coaches via intuitive and interpretive feedback in real-time. Our evaluation with actual coaching scenes confirms that REsCUE provides the informative cues to understand internal states of coachees. Since REsCUE is based on the unsupervised method and does not assume any prior knowledge, further applications beside executive coaching are conceivable using our framework.

## CCS CONCEPTS

• **Human-centered computing** → **Computer supported cooperative work**; *HCI design and evaluation methods*; • **Information systems** → Multimedia and multimodal retrieval;

---
[*]These authors contributed equally and are ordered alphabetically
[†]Also with University of Tsukuba, Japan.

## KEYWORDS

Executive Coaching, Nonverbal behavior analysis, Multimodal interaction, Anomaly detection

**Figure 1: REsCUE detects the behavioral cues of the coachee and notifies the coach in real-time to help the coach understand the internal states of the coachee.**

## 1 INTRODUCTION

*Executive coaching* plays an important role in human resource development [24, 35]. As a result, many companies invest in executive coaching to improve the leadership skills or the performances of their managers and the market share of executive coaching has increased to $2 billion [3, 25, 29].

Executive coaching usually consists of personal, one-on-one sessions [38, 55]. One-on-one sessions are preferred because coaches are required not only to build a rapport with a coachee but also to observe the nonverbal behavior of the coachee during the coaching session [6, 39]. For example, the

use of disorienting dilemmas is one of an important coaching process [12]; however, sensitive conversations about a coachee's dilemmas may cause deceptive responses [41].

In such a situation, coaches are expected to notice a discrepancy between the verbal response and the actual thoughts of the coachee using nonverbal cues [26]. Many articles list the observation skills as one of the skills required for effective coaching, in addition to emotional intelligence and questioning techniques [22, 27].

However, maintaining such objective observations throughout the coaching session requires a great deal of skill [49]. Coaches are often immersed in the verbal communication, paying attention to a deeper or emotional topic or thinking about what to ask next. In addition, self-deception may interfere with the quality of perception, e.g., ignoring significant behavior unconsciously based on faulty thinking or irrational beliefs [4]. Therefore, we expect that the quality of coaching could be improved if the coaches are automatically notified of important nonverbal behavioral cues from the coachee independently of their subjectivity or mental load.

One possible solution is to apply conventional methods proposed in the context of human activity analysis [1] or social signal processing [59]. However, these methods are mainly targeted at classifying human activities into specific categories and therefore have low affinity to the current situation, i.e., each behavior may correspond to various meanings depending on its context in the coaching session [8, 33]. In addition, most of these methods are designed for post-analysis and are not applicable for providing real-time feedback during a session. We, therefore, propose *REsCUE*, a new system introducing a real-time anomaly detection method into human behavior analysis (Figure 1).

Our framework, by exploiting the anomaly detection method, does not require prior knowledge or heuristic rules and therefore leaves room for the coach's interpretation of the semantics of the behavior based on the context. Moreover, combined with state-of-the-art behavior analysis methods, our framework is able to detect small but important behavioral cues, which might be missed by the coaches. REsCUE presents a new perspective of human behavior analysis that augments the perception of the user while leaving its interpretation to the user, which pave the way for further applications in the HCI community.

**Contribution**

The following four points are the main contributions of this study.

(1) We developed an intelligent system for use in coaching sessions that can automatically detect the nonverbal behavioral cues of coachees and provide feedback to coaches in real-time.

(2) Based on a preliminary analysis, we confirmed that the combination of the posture and gaze information is an effective modality for detecting nonverbal cues of coachees.

(3) Our user study in actual coaching scenes demonstrates that the proposed system provides informative feedback to professional coaches and would likely improve the quality of sessions.

(4) Because the proposed method is based on an unsupervised algorithm and does not assume prior knowledge, it can be applied widely outside of executive coaching applications, as a new framework for real-time behavioral analysis.

## 2 LITERATURE REVIEW

### Background

The relation between nonverbal behaviors and internal states has been a distinguished topic in the history of science [2, 30]. Beginning with Charles Darwin [18], many researchers have pointed out that nonverbal behaviors are spontaneous and unregulated expressions of internal states [19]. On the contrary, the effect of nonverbal behaviors on the internal states has also been revealed, e.g., the influence of facial muscular activity on people's affective responses [10].

Based on the relationship, observations of the nonverbal behaviors have been largely focused in various areas including executive coaching, as discussed in the "Introduction" section. For example, teachers are encouraged to pay attention to nonverbal behaviors of students, which convey their underlying feelings [36]. In addition, not only teachers or therapists [37] but also salespeople [48] or entrepreneurs [47] are expected to get a handle on nonverbal behavioral cues. Subsequently, a research domain of automatically analyzing nonverbal behaviors, which is often referred to as social signal processing [59], has spread.

### Related Work

Many methods have been proposed to analyze human nonverbal behavior using various modalities for one-on-one sessions or group discussions in the context of both human activity analysis [1] and social signal processing [59]. For example, conventional methods relying on handcrafted features have been widely researched, e.g., facial expression recognition [42, 53] and posture estimation [13]. Conversely, due to the development of deep learning in recent years, end-to-end methods using neural networks have become popular and have shown overwhelming performance improvements. For example, Wei et al. [61] achieved state-of-the-art performance in posture estimation by introducing a convolutional neural network.

Based on such analysis technologies, many applications have been proposed [59]. For example, Sanchez-Cortes et al. [51] proposed a method to detect emergent leaders in a group discussion using handcrafted audio and visual features. Beyan et al. [5] applied multiple kernel learning to similar features to predict the leadership styles of emergent leaders. Hoque et al. [32] leveraged multimodal behavioral data to generate instructive feedback in the context of training for job interviews. Nihei et al. [45] introduced a convolutional neural network to extract important utterances from multimodal behavioral data without relying on handcrafted features. However, these methods are designed to analyze sessions after they occur and are not formulated to provide feedback to coaches in real-time.

Based on the methods to understand human nonverbal behavior in real-time, some studies have proposed systems to provide real-time feedback on social interactions [17, 40, 43, 44, 52, 56, 58]. For example, Rhema [56] is designed to help people with public presentations by providing feedback in real-time via Google Glass based on a speaker's volume and speaking rate. Logue [17] addressed a similar situation by providing feedback via head-mounted display based on body energy and openness calculated from hand positions. For group discussions, Tausczik et al. [58] proposed a system to analyze the communication patterns of participants and to provide linguistic feedback for improving teamwork. In addition, Damian et al. [16] proposed a general framework to provide real-time feedback on predefined behavioral events represented in an XML format.

These methods are designed to provide explicit feedback based on some specific rules, such as "louder" if the speaking voice is faint or "pay attention to what others are saying" if the group dynamics are poor. However, as mentioned in the "Introduction" section, the meaning of the behavior is largely dependent on the context in a coaching session and therefore such explicit feedback would be impossible.

A similar discussion was presented in a proposal of AutoManner [57], a system to improve body languages in public speaking, that is "the appropriateness of the body language is largely dependent on the context of the speech—which is difficult to automatically assess." The system overcame this problem by displaying the estimated body skeleton with its changes and distributions in the time series and leaving room for interpretation by the speaker. However, it is specifically designed for public speaking and therefore not directly applicable to coaching. Moreover, it assumes the use of post-analysis by speakers themselves and therefore is not able to provide real-time feedback.

## 3 PROPOSED METHOD

In this section, we first describe the requirements of the proposed method. Then, the technical details of the method are presented, including how these requirements are solved.

### Requirements

To make the coaching sessions more effective using behavior analysis, the following requirements should be considered:

(1) **Unsupervised detection**
As discussed in the "Introduction" section, coaches are required to maintain objective and unbiased observations of the coachees. That is, assessing the behavior of the coachees based on heuristic rules or human-annotated training data introduces certain criteria and is not appropriate. In addition, due to the dependency of the meaning of nonverbal behaviors on their context, designing effective rules or collecting reliable training data is unrealistic. Therefore, we need the proposed system to the detect behavioral cues using an unsupervised algorithm.

(2) **Real-time feedback**
We aim to provide cues for coaches to understand the state of coachees to improve the quality of coaching sessions. Therefore, we need the proposed system to detect the cues and provide feedback in real-time, not via post-analysis of a session.

(3) **Intuitive and interpretive feedback**
We assume that coaches will use the proposed system while conversing with coachees. Therefore, the feedback presented to the coaches needs to be intuitive so that they can interpret it at a glance. At the same time, feedback that is too abstract, such as presenting only the fact that the behavior has changed or the statistical value of how it has changed, looses its context and is difficult for the coaches to interpret even though it would take a short time to understand. Therefore, we need the proposed system to preserve both intuitiveness and interpretiveness with regard to the feedback.

(4) **Non-interruptive notifications**
Similar to requirement (3), we need to consider how to notify the coaches of the feedback. If the notification causes an interruption, e.g., requesting an action by the coach every time, the quality of the coaching session may degrade. That is, we need the proposed system to notify in a non-interruptive manner.

(5) **Portable and non-interfering devices**
Because coaching sessions are often held in the office of the coachee, we need the proposed system to consist of portable devices so that the coach can easily transport them. In addition, if the proposed system requires the

coachee to wear devices or sensors, this may interfere with their concentration and result in an obstacle to building a rapport. Therefore, we also need the system to consist of non-interfering devices.

## Overview

To address the above requirements, we designed the proposed system as shown in Figure 2. The system collects the behavioral data of the coachee via external devices and obtains multimodal feature data. Using an anomaly detection algorithm, it detects important behavioral cues and notifies the coach both visually and tactilely in real-time.

We now describe the technical specifications of the proposed system along with the rationale behind specification.

## Multimodal Input

The proposed system obtains the multimodal behavioral data from the coachees to detect their behavioral cues.

As mentioned in the "Related Work" section, various types of multimodal features have been used in automated behavior analysis. For example, Nihei et al. [45] leveraged head pose information and speech features and concluded that the combination of these features achieved the best accuracy for detecting important utterances compared to unimodal methods. Beyan et al. [5] combined the pose and speech features with the gaze information and reported the effectiveness of the multimodal input. Hoque et al. [32] exploited facial expressions for the training of interviewees.

Based on both these studies and the coaching skills mentioned in the "Introduction" section, we prepared three input features: body posture (including head pose), gaze direction, and facial expression. In detail, along with the fact that the body language captured from the body posture is the basis of the observation skill described in [22, 27], the importance of interpreting the internal state from both the facial expression and the eye of the coachee are emphasized in [6]. Then, we selected which modalities to use in the proposed system later based on a preliminary experiment.

Here, we excluded the speech features due to requirement (4) because it is difficult to construct intuitive and interpretive feedback from changes in speech features. While presenting the frequency or the decibel of the coachee's voice is possible, it is difficult to understand how the behavior of the coachee changed and to infer their internal state from such feedback.

In addition, it is possible to use biometric devices to directly capture the signals from the coachee's body. However, this contradicts requirement (5) and could create a psychological barrier. Therefore, in this study, we limited the input modalities to those that are measurable without a body-mounted sensor.

*Posture.* The proposed system uses the pose estimation algorithm proposed in [61]. Because the algorithm is able to estimate the pose from images taken by a web camera, motion-tracking devices, which contradict requirement (5), are not required. In addition, the algorithm can not only achieve the state-of-the-art performance, as mentioned in the "Related Work" section, but can also be processed quickly enough to provide real-time feedback, satisfying requirement (2).

In the proposed system, the 12 key points shown in Figure 3 are used to detect the behavioral cues. We excluded the key points in the lower body because the coaching sessions are usually held at a desk.

*Gaze.* The proposed system uses a commercial eye tracker Tobii 4C to detect the gaze direction. This is a USB-connected device that is capable of tracking the looking direction without being worn, satisfying requirement (5).

In the proposed system, we used the two-dimensional coordinate of the gaze position. Here because the proposed system focuses on the relative change in the detected value, not the absolute value, our system does not require a calibration step for each session.

*Facial Expression.* The proposed system uses MicroExpNet [14] to extract the facial expressions of the coachees. This is a small and fast convolutional neural network designed for facial expression recognition, which is obtained by distilling a heavy and accurate neural network. Çugu et al. [14] reported that the network achieved over 95.0% classification accuracy for the eight expressions of "neutral," "anger," "contempt," "disgust," "fear," "happy," "sadness," and "surprise" under the real-time conditions. Therefore, we decided to use this network to meet requirement (2).

In the proposed system, we use the output value of the final layer after the softmax activation as an eight-dimensional feature vector of the facial expression in the same manner as [60]. In other words, each value of the vector represents the probability that the expression belongs to the corresponding class.

## Anomaly Detection

The proposed system detects behavioral cues from the multimodal feature data obtained by the method described in the "Multimodal Input" section. To satisfy requirements (1) and (2), we used anomaly detection algorithms, which are sometimes referred to as change point detection algorithms. This is because there are many proposed unsupervised online algorithms for anomaly detection [11].

We use SmartSifter [62], an adaptive anomaly detection algorithm based on the Gaussian Mixture Model (GMM). The main reason we chose the algorithm is that it is one of the most popular unsupervised online anomaly detection algorithms available. In addition, the results obtained with this

Figure 2: Overview of REsCUE. REsCUE detects important behavioral cues of the coachee via multimodal feature data and notifies the coach both visually and tactilely.



Figure 3: A real example showing the selected key posture points. The yellow circles represent five key points from the head: the nose, both eyes, and both ears. The red circles represent seven key points from the body: the neck, both shoulders, both elbows, and both wrists.

GMM-based approach are useful for designing informative feedback, as described later in this section.

Each time new input data arrive, SmartSifter estimates the data's outlierness based on the likelihood calculated by the GMM and, at the same time, updates the parameters of the GMM to fit the input data at the same time. More formally, let $x^{(t)}$ be an input data and $c_i^{(t)}$, $\mu_i^{(t)}$, and $\Sigma_i^{(t)}$ be the weight, mean, and covariance of the $i$-th component of the $l$-components GMM, respectively, at time $t$. Then, the outlierness of $x^{(t)}$ is calculated as follows.

$$a^{(t)} = -\ln \Sigma_{i=0}^l c_i^{(t-1)} \mathcal{N}\left(x^{(t)} \mid \mu_i^{(t-1)}, \Sigma_i^{(t-1)}\right) \qquad (1)$$

The parameters of the GMM are updated subsequently as follows.

$$
\begin{aligned}
\gamma_i^{(t)} &= \frac{c_i^{(t-1)} \mathcal{N}\left(x^{(t)} \mid \mu_i^{(t-1)}, \Sigma_i^{(t-1)}\right)}{\Sigma_{j=0}^l c_j^{(t-1)} \mathcal{N}\left(x^{(t)} \mid \mu_j^{(t-1)}, \Sigma_j^{(t-1)}\right)} \\
c_i^{(t)} &= (1-r)c_i^{(t-1)} + r\gamma_i^{(t)} \\
\bar{\mu}_i^{(t)} &= (1-r)\bar{\mu}_i^{(t-1)} + r\gamma_i^{(t)} x^{(t)} \qquad (2) \\
\mu_i^{(t)} &= \bar{\mu}_i^{(t)}/c_i^{(t)} \\
\bar{\Sigma}_i^{(t)} &= (1-r)\bar{\Sigma}_i^{(t-1)} + r\gamma_i^{(t)} x^{(t)} x^{(t)^T} \\
\Sigma_i^{(t)} &= \bar{\Sigma}_i^{(t)}/c_i^{(t)} - \mu_i^{(t)} \mu_i^{(t)^T}
\end{aligned}
$$

Here, $r$ represents a forgetting rate, which is related to the degree of discounting of past input data.

In the proposed system, SmartSifter is extended to use batches in the time series and therefore takes $X^{(t)} \in \mathbb{R}^{N \times M}$ as input, where $N$ represents the number of frames in a single batch and $M$ represents the number of the dimensions of the modality data. This is because the frame-by-frame behavioral changes would include instantaneous physiological responses, which make the feedback to the coach noisy. Introducing batch processing enables the proposed system to detect the changes in the distribution of the behavior. This increases the chance of capturing relatively long-term behavioral changes, which may be more difficult to recognize for a human observer [46, 54].

Moreover, the proposed system exploits SmartSifter to obtain the interpretive feedback, fulfilling requirement (3). Given that the GMM is used for clustering of the behavioral data, each component of the obtained GMM can be regarded as a representation of a particular behavioral state of the coachee. Therefore, the proposed system can obtain the representative $l$ frames in a batch at time $t$ as follows.

$$\hat{X}_i^{(t)} = X_{\hat{n}}^{(t)} \ (i = 1, \ldots, l) \qquad (3)$$
$$where \ \hat{n} = \underset{1 \le n \le N}{\operatorname{argmax}} \ \mathcal{N}\left(X_n^{(t)} \mid \mu_i^{(t-1)}, \Sigma_i^{(t-1)}\right)$$

Conversely, the most significant outlier frame can be obtained as follows.

$$\check{X}^{(t)} = X_{\check{n}}^{(t)} \qquad (4)$$
$$where \ \check{n} = \underset{1 \le n \le N}{\operatorname{argmin}} \ \Sigma_{i=0}^l c_i^{(t-1)} \mathcal{N}\left(X_n^{(t)} \mid \mu_i^{(t-1)}, \Sigma_i^{(t-1)}\right)$$

By observing both the representative frames from the last batch and the outlier frame from the current batch, the coach can easily understand how the behavior of the coachee has changed while preserving the interpretiveness of the change.

We summarize the above in Algorithm 1. Every time new input data arrive, the outlierness is calculated in the same manner as in Equation 1. If the outlierness exceeds the given threshold, the frames obtained using Equation 3, which show the representative states so far, and the current outlier frame obtained using Equation 4 are presented to the coach as the

**Algorithm 1:** The anomaly detection procedure

**Input:**

$l$: the number of components in GMM

$r$: the forgetting rate

$a_{th}$: the threshold of the outlierness to give feedback

**begin**

initialize $c_i^{(0)}, \boldsymbol{\mu}_i^{(0)}, \bar{\boldsymbol{\mu}}_i^{(0)}, \Sigma_i^{(0)}, \bar{\Sigma}_i^{(0)}$ $(i = 1, \ldots, l)$

$t \leftarrow 1$

**while** *the new input data $X^{(t)}$ is available* **do**

calculate the outlierness $a^{(t)}$ based on Eq. 1

**if** $t > 1$ **and** $a^{(t)} > a_{th}$ **then**

get the representative frames $\{\hat{X}_i^{(t-1)}\}$ (Eq. 3)

get the outlier frame $\check{X}^{(t)}$ (Eq. 4)

give feedback with $\{\hat{X}_i^{(t-1)}\}$ and $\check{X}^{(t)}$

**end**

update $c_i^{(t)}, \boldsymbol{\mu}_i^{(t)}, \bar{\boldsymbol{\mu}}_i^{(t)}, \Sigma_i^{(t)}, \bar{\Sigma}_i^{(t)}$ based on Eq. 2

$t \leftarrow t + 1$

**end**

**end**

feedback. Then, the parameters of the GMM are updated in the same way as in Equation 2.

**Feedback**

To satisfy requirement (4), we make use of tactile feedback to notify the coaches when behavioral cues are detected. In particular, the proposed system provides coaches with a smartwatch, which vibrates on detection of a behavioral cue. Tactile feedback is used because it does not interfere with the performance of concurrent tasks while obtaining high notice rates [20]. Moreover, the capability of tactile feedback during social interactions has been confirmed [15].

At the same time, representative frames from past scenes and the outlier frame from the current scene are displayed to coaches on detection of a behavioral cue. This allows coaches to easily understand how the behavior has changed, and this information is used to further infer and analyze the internal state of the coachee. In other words, this visualization achieves both the intuitiveness and interpretiveness as stipulated by requirement (3), because it enables the coaches to grasp feedback at a glance while making the feedback informative.

Moreover, because the scale of body movement patterns varies between individuals [21], it is better to provide the coaches with the capability to adjust the detection threshold during the sessions. Therefore, we placed "more" and "less" buttons on the smartwatch to change the threshold $a_{th}$ in

Algorithm 1. In this way, the coaches are able to control the frequency of notifications according to each coachee.

In summary, combining the tactile notification via a smartwatch with intuitive and interpretive visual information, coaches are freed from the burden of paying their attention to displays in parallel to conversing with the coachees. Moreover, the coaches can easily adjust the sensitivity of the detection on the smartwatch depending on the characteristics of each coachee. Therefore, the coaches can exploit feedback without losing concentration and the better coaching performances are expected.

## 4 PRELIMINARY EXPERIMENT

To determine which modality to use in the proposed system, we conducted a preliminary experiment. In this section, we describe the procedure and results of the experiment.

**Data Collection**

The experiment involved three professional coaches (aged 25–39 years old), who participated voluntarily. Each coach had coaching sessions with two different coachees for at least 30 minutes (4h 28m 35s in total). All participants, including both the coaches and the coachees, agreed to the use of the collected data for the research purposes.

During the sessions, the behavior of the coachees was recorded with a video camera and a Tobii eye tracker. After each session, the participating coaches are asked to watch the recorded video and list the top 10 most important behavioral cues of the coachee to infer their internal states.

**Implementation**

We implemented the proposed algorithm and applied it to the recorded data. Based on our empirical observations, the number of components and the forgetting rate in Algorithm 1 were set to 2 and 0.1, respectively. In detail, we found that, as the number of components increased, not only did it take more time until the model's initial convergence, but also it became more difficult for the coach to interpret the detected cues since a larger number of frames were displayed simultaneously. To compare the detected results with the behavioral cues pointed out by the coaches, we had the system to output the top 10 most significant peaks in the outlierness instead of specifying the threshold. Here, the first three minutes of each session are excluded for the detection because it takes several minutes for the parameters of the GMM to converge, as shown in Figure 4.

In addition, we had the proposed system to obtain new behavioral data on every 0.5 seconds and combine them into 30-seconds batches to ensure that the same processing performance was reproducible in a portable computing environment, e.g., a regular laptop with a single GPU, as stipulated in requirement (5). This is also expected to make the

**Figure 4: A real example of the transition of the calculated outlierness. The orange points show the detected cues, which are the top 10 peaks of the outlierness.**

**Table 1: The results of the preliminary experiment. The combination of the posture and gaze information showed the highest detection performance.**

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

feedback less noisy, as discussed in the "Anomaly Detection" section.

### Evaluation Metrics

To compare the effectiveness of each modality, we chose the recall and the minimizing Kendall's $\tau$ distance [23] as evaluation metrics. The recall represents how many behavioral cues, which are pointed out by the coaches, are captured by the proposed system. In this case, we set the error tolerance to 30 seconds. This is because, in addition to the constraint of the batch size, some behavioral cues, such as stretching or scratching one's head, take more than several seconds making it difficult to specify their precise timing.

The minimizing Kendall's $\tau$ distance is a metric to measure the similarity between two top $k$ lists ($k \geq 2$) and is widely used in the evaluation of search engines [31, 64]. Given the two lists $\tau_1$ and $\tau_2$, the distance is defined as follows.

$$K_{min}\left(\tau_1, \tau_2\right) = \frac{\Sigma_{\{i,j\} \in P(\tau_1, \tau_2)} \bar{K}_{i,j}\left(\tau_1, \tau_2\right)}{k \times (k-1)/2}$$

Here, $P\left(\tau_1, \tau_2\right)$ denotes the set of all unordered pairs of distinct elements in $\tau_1 \cup \tau_2$. Then, $\bar{K}_{i,j}\left(\tau_1, \tau_2\right) = 1$ if (i) $i$ appears only in one list and $j$ appears only in the other list, (ii) $i \prec j$ in one list and only $j$ appears in the other list, or (iii) $i \prec j$ in one list and $i \succ j$ in the other list; otherwise, $\bar{K}_{i,j}\left(\tau_1, \tau_2\right) = 0$. Consequently, if $\tau_1$ and $\tau_2$ are identical, $K_{min}\left(\tau_1, \tau_2\right) = 0$.

### Results

The results are shown in Table 1. From the comparison of the recall and the minimizing Kendall's $\tau$ distance, we confirmed that the combination of the posture and gaze information is the most suitable for detecting behavioral cues in coaching sessions. In addition, as discussed in the "Multimodal Input" section, the experiment demonstrated the effectiveness of the multimodal features compared to the unimodal features

by observing the difference with cases of only the posture or the gaze was used.

The facial expression information, however, did not contribute to improvements in the detection performance. Here, the chance rate of the recall is 0.11, meaning that, if we choose the cue points randomly, 1 of the 10 chosen points is considered correct on average. However, the recall of the case using only the facial expression is lower than that.

### Analysis

In this subsection, we discuss the reasons behind the results in Table 1 by analyzing the detected cues according to each modality.

*Why was the combination of the posture and gaze information effective?* In the recorded data, based on only the posture information, the proposed system detected a wide range of behavioral cues ranging from leaning on a chair to putting a hand on one's hip[1] or placing a hand on the back of one's neck[2], as shown in Figure 5. Likewise, the gaze information detected not only changes in the looking direction but also *self-touch* cues [9] such as rubbing one's eyebrow[3]. This is because self-touch cues often interfere with the detection of the eyes and be captured by the anomaly detection algorithm. Therefore, the effectiveness of the combination of the posture and gaze information can be attributed to the capability of the system to detect a variety of the behavioral cues.

*Why was the facial expression ineffective?* As previously stated in the "Results" section, the facial expression did not improve the accuracy of detecting behavioral cues in the sessions.

---

[1]Putting a hand on one's hip is considered to represent a defensive state [63].

[2]Placing a hand on the back of one's neck is considered to represent an aggressive state [28].

[3]Rubbing an eyebrow is considered to represent an anxious state [9].

| Used modalities | Detected behaviors | Detected outlier frames | Precedent representative frames |
|---|---|---|---|
| | Leaning on a chair | | |
| Posture information | Putting a hand on one's hip | | |
| | Placing a hand on the back of one's neck | | |
| Gaze information | Looking at different directions from the coach | | |
| | Rubbing one's eyebrow | | |

**Figure 5: A real example of detected behavioral cues from the one of the recorded sessions. The top three cues are detected using only the posture information, and the next two cues are detected using only with the gaze information. Compared to the preceding representative frames, the changes can easily be seen.**

There appear to be two reasons for this result. First, facial expressions are obvious and superficial; therefore, the coaches do not regard them as important behavioral cues reflecting the internal states of coachees. Second, even though the neural network used in the experiment is state-of-the-art, the accuracy might not be sufficiently high. This is attributable to the fact that MicroExpNet is designed not for faces in free conversation but for posed faces. Therefore, for example, a face with an open mouth may be classified as fearful even though the coachee is just talking.

*What was the difference between the coaches and the proposed system?* From the results, at least 30% of the detected points were not included in the behavioral cues listed by the coaches. When we showed such points to the participating coaches, the points were roughly divided into two groups based on their responses. The first was a group of obvious and non-informative points such as opening one's notebook or sneezing. It is because the proposed method uses an unsupervised algorithm and cannot take the meanings of the detected points into consideration. This result suggests the importance of designing non-interruptive notification, as in



**Figure 6: The interface displaying the feedback. The left half presents representative frames from past scenes, and the right half presents the top two outlier frames from current scenes.**

requirement (4), so that the coach can easily ignore feedback when it is non-informative.

The other group included points that the coaches agreed were informative. One participant said: "Although I did not notice this when I watched the recorded video, once the system pointed out that the coachee had bent slightly forward, I could see that he was opening his mind from about that moment." This comment suggests that the proposed system could contribute to improving the quality of coaching sessions by providing the feedback in real-time.

## 5 USER STUDY

To confirm the effectiveness of the proposed system, we conducted a user study. In this section, we describe its setting and results.

### Implementation

We implemented a complete version of the proposed system to perform a user study. The same parameters as the "Data Collection" section were used except for the detection threshold. The system uses the outlierness of the first peak after three minutes from the beginning of each session as the initial threshold and allows coaches to adjust the threshold subsequently via their smartwatches, as mentioned in the "Feedback" section.

The feedback indicating the behavioral cues is presented in Figure 6. It displays the representative frames from past scenes and outlier frames from current scenes so that the coach can understand how the behavior of the coachee has changed. At the same time, the coach's smartwatch vibrates and shows buttons to adjust the threshold as shown in Figure 7. Pressing the "more" button decreases the threshold while pressing the "less" button increases the threshold.

**Figure 7: The interface of the control buttons shown on the smartwatch. The coach can adjust the threshold for detecting behavioral cues by pressing these buttons.**



**Figure 8: An example of the setup for the user study. The behaviors of the participating coachees were tracked using a web camera and an eye tracker. The participating coaches used a laptop and a smartwatch to receive the feedback.**

### Procedure

In the user study, five professional coaches (aged 25–39 years old), including the three coaches from the preliminary experiment, participated voluntarily. Each coach had coaching sessions with three different coachees (15 sessions in total) using the proposed system, as shown in Figure 8.

Then, we conducted short interviews with the participating coaches. In this interview, we asked for their subjective opinions concerning the usability of the proposed system. We also asked whether the given feedback effectively improved the quality of the sessions. If the participant agreed on the effectiveness, we asked how the sessions changed due to the feedback.

### Comments

We found all of the participating coaches responded positively about the proposed system in the subjective interview. Here, we separately examine the obtained comments concerning the usability and the effectiveness.

*The usability of the proposed system.* When we asked about the usability of the proposed system, the replies were affirmative, such as "There was nothing confusing or difficult to understand." and "It was so easy to use that I can imagine that I am using it from tomorrow."

More specifically, concerning the visualization of the feedback, one participant responded,

> "Putting the past frames side by side makes the changes in the behavior obvious."

Another participant commented on the comparison with the explicit feedback:

> "Simpler feedback such as just showing 'defensive' or 'opening one's heart' could also be easy to understand. However, if it contradicts my assumptions, 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 while reminding me of other possibilities."

Concerning the tactile notification, one participant responded,

> "I think the notification is very good because it does not break my concentration and it is not noticed by the subject."

From a different perspective, another participant commented on the benefit of the tactile notification:

> "When having sessions with about seven or eight clients a day, I sometimes feel out of it. At such a time, the tactile feedback would help me focus on the sessions."

The above comments support the usability of the proposed system, as well as the suitability of the feedback design.

At the same time, one participant gave us suggestions for future improvements. He suggested visualizing the trends of the behavioral cues throughout the session, or throughout multiple sessions of the same coachee:

> "Further inferences are possible if this shows that the coachee repeats similar behaviors or that the trend in the behavioral cues changes depending on the topic of conversation."

This can be accomplished by applying a clustering algorithm in an unsupervised manner. For example, the algorithm enables the similarity with past scenes to be represented by visualizing the cue that each cluster belongs to in a time series.

Moreover, this could lead the proposed system to determine the non-informative behavioral cues. In particular, by adding an "obvious" button to the smartwatch in the same manner as shown in Figure 7, clusters of the non-informative cues could be identified in an interactive manner. In this way,

the presence of obvious and non-informative cues, which were discussed in the "What was the difference between the coaches and the proposed system?" section, can be reduced. Therefore, we would like to implement this feature in the near future.

*The effectiveness of the proposed system to improve the quality of the sessions.* The participating coaches also commented positively on the informativeness of the detected behavioral cues and the effectiveness of the proposed system in the sessions, for example:

> "Although I often immerse myself in the conversation, thanks to this system, I was able to pay attention to the behavior of the client."

and

> "This system made me realize that I unconsciously missed or ignored many important behavioral cues."

Other comments confirmed that the proposed system helped the coaches change the content of the sessions in accordance with the state of the coachees:

> "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."

In addition, one participant commented on the educational aspect of the proposed system:

> "Up to this time, to cultivate observational skills, we had to review the recordings of the sessions of ourselves or observe the sessions by other coaches. However, using this system, it would be possible to learn what sort of behavioral cues should be focused on during a session and reduce the training time."

The above comments suggest that the proposed system may effectively improve the quality of the coaching sessions.

## 6 DISCUSSION

Although the proposed system was generally appreciated in the "User Study" section, there is still room for further exploration. In this section, we discuss the limitations and future directions of our research.

### Limitations

Throughout the preliminary study, the combination of the posture and gaze information was confirmed to be the most effective and thus chosen as the input modalities in the following user study. Nonetheless, the number of the participants was relatively small to rule out other possibilities. Additional investigations with other available modalities are desirable to seek for potential combinations.

Also, though the effectiveness of the proposed system is qualitatively supported by the comments from the participated coaches in the user study, a quantitative comparison of the outcome of the coaching sessions with controlled groups is preferred so as to avoid subject biases. However, the impact of executive coaching is shaped by a variety of factors such as its purpose, length, organizational context, and individual differences [34] and evaluating its outcome via randomized controlled experiments is costly [3]. One possible remedy is to expand the preliminary experiment to support the results from the other aspects, e.g., collecting self-labelled ground-truth data of the internal states from coachees to validate whether the change of their internal states is captured using the proposed system.

### Future Directions

Throughout the user study, the effectiveness of the real-time feedback is confirmed. In particular, changing the direction of the session on the spot based on the detected cues is not achievable using the conventional methods designed for post-analysis, as we mentioned in the "Related Work" section.

On the other hand, though the design of the proposed feedback system is based on the rationale presented in the "Requirements" section, there are other possibilities like those proposed by previous studies. We would like to explore a better interaction with the coaches such as comparing different ways of presenting cues, or suppressing non-informative notifications using a clustering method, which is discussed in the "The usability of the proposed system" section.

Exploring cases of further use also remains a promising endeavor. Since the proposed method consists of unsupervised learning and does not require any prior knowledge or rules, it could be used to analyze the behavior of people outside coaching sessions. For example, REsCUE might be able to assist people working in dementia care, where it is necessary to analyze the behavior of a patient and consider therapeutic approaches [7]. Moreover, the connection between conversation and structural neural connectivity in children has been elucidated recently [50], and thus REsCUE would potentially be utilized in early childhood education as well.

## 7 CONCLUSIONS

In this study, we introduced REsCUE, an intelligent system for use in coaching sessions that can automatically detect nonverbal behavioral cues of coachees and provide feedback to coaches in real-time. Based on a preliminary experiment, the posture and gaze information proved to be effective modalities to detect behavioral cues. In actual sessions with professional coaches, a number of favorable comments were obtained, indicating that REsCUE can help coaches to maintain a conversation with coachees while simultaneously inferring their internal states. For future work, we will investigate other applications of REsCUE by exploiting that the proposed method is based on the unsupervised algorithm and does not depend on prior information.

## REFERENCES

[1] Jagdishkumar Aggarwal and Michael S. Ryoo. 2011. Human activity analysis: A review. *Comput. Surveys* 43, 3 (2011), 16:1–16:43. https://doi.org/10.1145/1922649.1922653

[2] Stanley Feldstein Aron W. Siegman (Ed.). 1978. *Nonverbal Behavior and Communication.* Lawrence Erlbaum Associates, Hillsdale, NJ.

[3] Andromachi Athanasopoulou and Sue Dopson. 2018. A systematic review of executive coaching outcomes: Is it the journey or the destination that matters the most? *The Leadership Quarterly* 29, 1 (2018), 70–88. https://doi.org/10.1016/j.leaqua.2017.11.004

[4] Tatiana Bachkirova. 2014. Psychological development in adulthood and coaching. In *The Complete Handbook of Coaching* (2 ed.), Elaine Cox, Tatiana Bachkirova, and David Clutterbuck (Eds.). SAGE Publications, London.

[5] Cigdem Beyan, Francesca Capozzi, Cristina Becchio, and Vittorio Murino. 2018. Prediction of the Leadership Style of an Emergent Leader Using Audio and Visual Nonverbal Features. *IEEE Transaction on Multimedia* 20, 2 (2018), 441–456. https://doi.org/10.1109/TMM.2017.2740062

[6] Gary S. Bloom, Claire L. Castagna, Ellen Moir, and Betsy Warren (Eds.). 2005. *Blended Coaching: Skills and Strategies to Support Principal Development.* Corwin Press, Thousand Oaks, CA.

[7] Dawn Brooker. 2003. What is person-centred care in dementia? *Reviews in Clinical Gerontology* 13, 3 (2003), 215–222. https://doi.org/10.1017/S095925980400108X

[8] Kirk W. Brown and D. S. Moskowitz. 1998. Dynamic Stability of Behavior: The Rhythms of Our Interpersonal Lives. *Journal of Personality* 66, 1 (1998), 105–134. https://doi.org/10.1111/1467-6494.00005

[9] Nathan David Butzen, Victor Bissonnette, and Dan McBrayer. 2005. Effects of Modeling and Topic Stimulus on Self-Referent Touching. *Perceptual and Motor Skills* 101, 2 (2005), 413–420. https://doi.org/10.2466/pms.101.2.413-420

[10] Laurie Carr, Marco Iacoboni, Marie-Charlotte Dubeau, John C. Mazziotta, and Gian Luigi Lenzi. 2003. Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences* 100, 9 (2003), 5497–5502. https://doi.org/10.1073/pnas.0935845100

[11] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. *Comput. Surveys* 41, 3 (2009), 15:1–15:58. https://doi.org/10.1145/1541880.1541882

[12] Elaine Cox. 2015. Coaching and Adult Learning: Theory and Practice. *New Directions for Adult and Continuing Education* 2015, 148 (2015), 27–38. https://doi.org/10.1002/ace.20149

[13] Rita Cucchiara, Costantino Grana, Andrea Prati, and Roberto Vezzani. 2005. Probabilistic posture classification for human-behavior analysis. *IEEE Transaction on Systems, Man, and Cybernetics, Part A* 35, 1 (2005), 42–54. https://doi.org/10.1109/TSMCA.2004.838501

[14] Ilke Çugu, Eren Sener, and Emre Akbas. 2017. MicroExpNet: An Extremely Small and Fast Model For Expression Recognition From Frontal Face Images. *arXiv* 1711.07011 (2017), 1–9. http://arxiv.org/abs/1711.07011

[15] Ionut Damian and Elisabeth André. 2016. Exploring the Potential of Realtime Haptic Feedback during Social Interactions. In *Proceedings of the 10th International Conference on Tangible and Embedded Interaction.* ACM, New York, NY, 410–416. https://doi.org/10.1145/2839462.2856519

[16] Ionut Damian, Tobias Baur, and Elisabeth André. 2016. Measuring the impact of multimodal behavioural feedback loops on social interactions. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction.* ACM, New York, NY, 201–208. https://doi.org/10.1145/2993148.2993174

[17] Ionut Damian, Chiew Seng Sean Tan, Tobias Baur, Johannes Schöning, Kris Luyten, and Elisabeth André. 2015. Augmenting Social Interactions: Realtime Behavioural Feedback using Social Signal Processing Techniques. In *Proceedings of the 2015 ACM SIGCHI Conference on Human Factors in Computing Systems.* ACM, New York, NY, 565–574. https://doi.org/10.1145/2702123.2702314

[18] Charles Darwin. 1872. *The Expression of the Emotions in Man and Animals.* John Murray, London.

[19] Bella M DePaulo. 1992. Nonverbal behavior and self-presentation. *Psychological bulletin* 111, 2 (1992), 203.

[20] Aaron E. Sklar and Nadine Sarter. 2000. Good Vibrations: Tactile Feedback in Support of Attention Allocation and Human-Automation Coordination in Event-Driven Domains. *Human Factors* 41 (2000), 543–52. https://doi.org/10.1518/001872099779656716

[21] Paul Ekman, Wallace V Friesen, Maureen O'Sullivan, and Klaus Scherer. 1980. Relative importance of face, body, and speech in judgments of personality and affect. *Journal of Personality and Social Psychology* 38, 2 (1980), 270–277. https://doi.org/10.1037/0022-3514.38.2.270

[22] Andrea D. Ellinger, Alexander E. Ellinger, and Scott B. Keller. 2003. Supervisory coaching behavior, employee satisfaction, and warehouse employee performance: A dyadic perspective in the distribution industry. *Human Resource Development Quarterly* 14, 4 (2003), 435–458. https://doi.org/10.1002/hrdq.1078

[23] Ronald Fagin, Ravi Kumar, and D. Sivakumar. 2003. Comparing Top k Lists. *SIAM Journal of Discrete Mathematics* 17, 1 (2003), 134–160. https://doi.org/10.1137/S0895480102412856

[24] Daniel C. Feldman and Melenie J. Lankau. 2005. Executive Coaching: A Review and Agenda for Future Research. *Journal of Management* 31, 6 (2005), 829–848. https://doi.org/10.1177/0149206305279599

[25] Annette Fillery-Travis and David Lane. 2006. Does coaching work or are we asking the wrong question? *International Coaching Psychology Review* 1, 1 (2006), 24–36.

[26] Carol Kinsey Goman. 2008. *The Nonverbal Advantage: Secrets and Science of Body Language at Work.* Berrett-Koehler, San Francisco, CA.

[27] Anthony M. Grant. 2007. Enhancing coaching skills and emotional intelligence through training. *Industrial and Commercial Training* 39, 5 (2007), 257–266. https://doi.org/10.1108/00197850710761945

[28] Ewan C. Grant. 1969. Human Facial Expression. *Man* 4, 4 (1969), 525–692. https://doi.org/10.2307/2798193

[29] Robert G. Hamlin, Andrea D. Ellinger, and Rona S. Beattie. 2008. The emergent 'coaching industry': a wake-up call for HRD professionals. *Human Resource Development International* 11, 3 (2008), 287–305. https://doi.org/10.1080/13678860802102534

[30] Robert Gale Harper, Joseph D Matarazzo, and Arthur N Wiens. 1978. *Nonverbal Communication : The State of the Art.* Wiley, New York, NY.

[31] Taher H. Haveliwala. 2003. Topic-Sensitive PageRank: A Context-Sensitive Ranking Algorithm for Web Search. *IEEE Transaction on Knowledge and Data Engineering* 15, 4 (2003), 784–796. https://doi.org/10.1109/TKDE.2003.1208999

[32] Mohammed (Ehsan) Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: my automated conversation coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, NY, 697–706. https://doi.org/10.1145/2493432.2493502

[33] Patrizia M Ianiro and Simone Kauffeld. 2014. Take care what you bring with you: How coaches' mood and interpersonal behavior affect coaching success. *Consulting Psychology Journal: Practice and Research* 66, 3 (2014), 231–257. https://doi.org/10.1037/cpb0000012

[34] Baek-Kyoo Joo. 2005. Executive coaching: A conceptual framework from an integrative review of practice and research. *Human Resource Development Review* 4, 4 (2005), 462–488. https://doi.org/10.1177/1534484305280866

[35] Richard R. Kilburg. 2000. *Executive coaching: Developing managerial wisdom in a world of chaos.* American Psychological Association, Washington, DC. https://doi.org/10.1037/10355-000

[36] Gail King. 1999. *Counselling skills for teachers: Talking matters.* Open University Press, Berkshire.

[37] Lance M Kinseth. 1989. Nonverbal training for psychotherapy: Overcoming barriers to clinical work with client nonverbal behavior. *The Clinical Supervisor* 7, 1 (1989), 5–25. https://doi.org/10.1300/J001v07n01_02

[38] Richard Koonce. 1994. One on One. *Training & Development* 48, 2 (1994), 34–40.

[39] Karren Kowalski and Colleen Casper. 2007. The Coaching Process: An Effective Tool for Professional Development. *Nursing Administration Quarterly* 31, 2 (2007), 171–179. https://doi.org/10.1097/01.NAQ.0000264867.73873.1a

[40] Kazutaka Kurihara, Masataka Goto, Jun Ogata, Yosuke Matsusaka, and Takeo Igarashi. 2007. Presentation sensei: a presentation training system using speech and image processing. In *Proceedings of the 9th ACM International Conference on Multimodal Interaction*. ACM, 358–365. https://doi.org/10.1145/1322192.1322256

[41] Paula V. Lippard. 1988. "Ask me no questions, I'll tell you no lies": Situational exigencies for interpersonal deception. *Western Journal of Speech Communication* 52, 1 (1988), 91–103. https://doi.org/10.1080/10570318809389627

[42] Michael J. Lyons, Shigeru Akamatsu, Miyuki Kamachi, and Jiro Gyoba. 1998. Coding Facial Expressions with Gabor Wavelets. In *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*. IEEE Computer Society, Washington, DC, 200–205. https://doi.org/10.1109/AFGR.1998.670949

[43] Skanda Muralidhar, Jean M. R. Costa, Laurent Son Nguyen, and Daniel Gatica-Perez. 2016. Dites-moi: wearable feedback on conversational behavior. In *Proceedings of The 15th International Conference on Mobile and Ubiquitous Multimedia*. ACM, New York, NY, 261–265. https://doi.org/10.1145/3012709.3012733

[44] Anh-Tuan Nguyen, Wei Chen, and Matthias Rauterberg. 2012. Online feedback system for public speakers. In *Proceedings of the 2012 IEEE Symposium on E-Learning, E-Management and E-Services*. IEEE Computer Society, Washington, DC, 1–5. https://doi.org/10.1109/IS3e.2012.6414963

[45] Fumio Nihei, Yukiko I. Nakano, and Yutaka Takase. 2017. Predicting meeting extracts in group discussions using multimodal convolutional neural networks. In *Proceedings of the 19th ACM International Conference on Multimodal Interaction*. ACM, New York, NY, 421–425. https://doi.org/10.1145/3136755.3136803

[46] Erik C Nook, Kristen A Lindquist, and Jamil Zaki. 2015. A new look at emotion perception: Concepts speed and shape facial emotion recognition. *Emotion* 15, 5 (2015), 569–578. https://doi.org/10.1037/a0039166

[47] Kęstutis Peleckis, Valentina Peleckienė, Kęstutis Peleckis, and Tatjana Polajeva. 2016. Towards sustainable entrepreneurship: role of nonverbal communication in business negotiations. *Entrepreneurship and Sustainability Issues* 4, 2 (2016), 228–239. https://doi.org/10.9770/jesi.2016.4.2(10)

[48] Joseph O. Rentz, C. David Shepherd, Armen Tashchian, Pratibha A. Dabholkar, and Robert T. Ladd. 2002. A measure of selling skill: Scale development and validation. *The Journal of Personal Selling & Sales Management* 22, 1 (2002), 13–21.

[49] Ronald E. Riggio and Joanne Lee. 2007. Emotional and interpersonal competencies and leader development. *Human Resource Management Review* 17, 4 (2007), 418–426. https://doi.org/10.1016/j.hrmr.2007.08.008

[50] Rachel R. Romeo, Joshua Segaran, Julia A. Leonard, Sydney T. Robinson, Martin R. West, Allyson P. Mackey, Anastasia Yendiki, Meredith L. Rowe, and John D.E. Gabrieli. 2018. Language Exposure Relates to Structural Neural Connectivity in Childhood. *The Journal of Neuroscience* 38, 36 (2018), 7870–7877. https://doi.org/10.1523/JNEUROSCI.0484-18.2018

[51] Dairazalia Sanchez-Cortes, Oya Aran, Marianne Schmid Mast, and Daniel Gatica-Perez. 2012. A Nonverbal Behavior Approach to Identify Emergent Leaders in Small Groups. *IEEE Transaction on Multimedia* 14, 3-2 (2012), 816–832. https://doi.org/10.1109/TMM.2011.2181941

[52] Jan Schneider, Dirk Börner, Peter van Rosmalen, and Marcus Specht. 2015. Presentation Trainer, your Public Speaking Multimodal Coach. In *Proceedings of the 17th ACM International Conference on Multimodal Interaction*. ACM, New York, NY, 539–546. https://doi.org/10.1145/2818346.2830603

[53] Caifeng Shan, Shaogang Gong, and Peter W. McOwan. 2009. Facial expression recognition based on Local Binary Patterns: A comprehensive study. *Image Vision Computing* 27, 6 (2009), 803–816. https://doi.org/10.1016/j.imavis.2008.08.005

[54] Gillian Slessor, Louise H Phillips, and Rebecca Bull. 2008. Age-related declines in basic social perception: evidence from tasks assessing eye-gaze processing. *Psychology and Aging* 23, 4 (2008), 812–822. https://doi.org/10.1037/a0014348

[55] Lewis R Stern. 2004. Executive coaching: A working definition. *Consulting Psychology Journal: Practice and Research* 56, 3 (2004), 154–162. https://doi.org/10.1037/1065-9293.56.3.154

[56] Mohammad Iftekhar Tanveer, Emy Lin, and Mohammed (Ehsan) Hoque. 2015. Rhema: A Real-Time In-Situ Intelligent Interface to Help People with Public Speaking. In *Proceedings of the 20th Annual Meeting of the Intelligent Interfaces*. ACM, New York, NY, 286–295. https://doi.org/10.1145/2678025.2701386

[57] Mohammad Iftekhar Tanveer, Ru Zhao, Kezhen Chen, Zoe Tiet, and Mohammed (Ehsan) Hoque. 2016. AutoManner: An Automated Interface for Making Public Speakers Aware of Their Mannerisms. In *Proceedings of the 21st Annual Meeting of the Intelligent Interfaces*. ACM, New York, NY, 385–396. https://doi.org/10.1145/2856767.2856785

[58] Yla R. Tausczik and James W. Pennebaker. 2013. Improving teamwork using real-time language feedback. In *Proceedings of the 2013 ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, 459–468. https://doi.org/10.1145/2470654.2470720

[59] Alessandro Vinciarelli, Maja Pantic, Dirk Heylen, Catherine Pelachaud, Isabella Poggi, Francesca D'Errico, and Marc Schröder. 2012. Bridging the Gap between Social Animal and Unsocial Machine: A Survey of Social Signal Processing. *IEEE Transaction on Affective Computing* 3, 1 (2012), 69–87. https://doi.org/10.1109/T-AFFC.2011.27

[60] Duc Minh Vo and Thai Hoang Le. 2016. Deep generic features and SVM for facial expression recognition. In *Proceedings of the 3rd National*

*Foundation for Science and Technology Development Conference on Information and Computer Science.* IEEE Computer Society, Washington, DC, 80–84. https://doi.org/10.1109/NICS.2016.7725672

[61] Shih-En Wei, Varun Ramakrishna, Takeo Kanade, and Yaser Sheikh. 2016. Convolutional Pose Machines. In *Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition.* IEEE Computer Society, Washington, DC, 4724–4732. https://doi.org/10.1109/CVPR.2016.511

[62] Kenji Yamanishi, Jun'ichi Takeuchi, Graham J. Williams, and Peter Milne. 2004. On-Line Unsupervised Outlier Detection Using Finite Mixtures with Discounting Learning Algorithms. *Data Mining and Knowledge Discovery* 8, 3 (2004), 275–300. https://doi.org/10.1023/B:DAMI.0000023676.72185.7c

[63] Li Zhang and Bryan Yap. 2012. Affect Detection from Text-Based Virtual Improvisation and Emotional Gesture Recognition. *Advances in Human-Computer Interaction* 2012, 461247 (2012), 1–12. https://doi.org/10.1155/2012/461247

[64] Cai-Nicolas Ziegler, Sean M. McNee, Joseph A. Konstan, and Georg Lausen. 2005. Improving recommendation lists through topic diversification. In *Proceedings of the 14th International Conference on World Wide Web.* ACM, New York, NY, 22–32. https://doi.org/10.1145/1060745.1060754