Proceedings of the 2025 SICE Festival with Annual Conference
September 9-12, 2025, Chiang Mai, Thailand

Proposal of Methods and Approaches to Analyze the Impact of Daily Conversations on Individual Performance

Kano Hidetaka[1†] and Manabu Ichikawa[2] and Kaede Hujita[3]

[1]Systems Engineering and Science, Graduate School of Engineering and Science, University of Shibaura Institute of Technology, Tokyo, Japan
(Tel: +81-70-1569-5965; E-mail: br21088@shibaura-it.ac.jp)
[2]College of Systems Engineering and Science, University of Shibaura Institute of Technology, Tokyo, Japan
(Tel: +81-48-720-623; E-mail: m-ichi@sic.shibaura-it.ac.jp)
[3]Functional Control Systems, Graduate School of Engineering and Science,University of Shibaura Institute of Technology, Tokyo, Japan
(Tel: +81-48-720-623; E-mail: nb23113@sic.shibaura-it.ac.jp)

Abstract: Many organizations have tried to improve communication through various efforts, but most focused on settings like meetings or phone calls. Past studies mainly analyzed verbal interactions, with few examining nonverbal cues such as facial expressions and gestures. This study investigates how nonverbal communication in everyday lab conversations relates to productivity. Using Raspberry Pi devices to record and analyze interactions, the results show that serious expressions during discovery or learning promote task progress. Stable emotions made participants more open to advice, while emotional shifts often led to helpful support for completing tasks.

Keywords: nonverbal communication, sensors, daily conversations, performance

1. INTRODUCTION

1.1. Background

With the diversification of work styles—accelerated by the COVID-19 pandemic—communication in workplaces has declined, weakening human relationships and contributing to mental health concerns[1][2][3].

In response, some organizations have begun analyzing communication quality and emotional states using data from phone calls and meetings, but few studies focus on everyday, informal conversation[4][5].

1.2. Previous Research

Prior studies fall into three categories: (1) visualizing communication networks, (2) identifying key communication features, and (3) proposing analytical technologies. Examples include interaction analysis via wearable sensors (Abumi, 2023)[6], probabilistic modeling of chat data (Nonaka et al., 2022)[7], and stress detection with biosensors (Tomita et al., 2017)[8]. While these studies target verbal or spatial communication, few address universally visible nonverbal cues like gaze and facial expressions. Moreover, sensor-based methods often impose burdens on participants.

1.3. Purpose of the Study

This study aims to explore the relationship between language-independent nonverbal behavior in casual conversations and organizational productivity. By focusing on body-based nonverbal cues using non-contact video analysis, it seeks to demonstrate the overlooked value of daily communication and support broader methods for organizational communication assessment.

† Hidetaka Kano is the presenter of this paper.

2. METHOD

This study, conducted in a university lab, analyzed video-recorded daily conversations of undergraduate and graduate students to examine the relationship between nonverbal communication and weekly goal achievement. The initial experiment assessed how emotions, gestures, and face orientation related to task progress. The additional experiment revised some indicators and included new participants to improve classification accuracy, using the same analysis framework.

3. VIDEO RECORDING

3.1. Equipment Used

Video recording in the laboratory was performed using a Raspberry Pi 5[9], a single-board computer, connected to the official camera module (Raspberry Pi Camera Module 2)(Fig. 1).

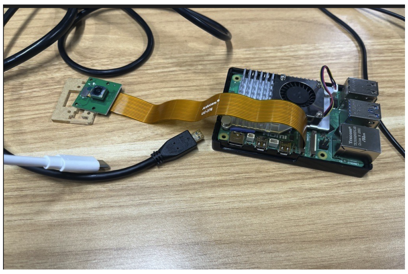

Fig. 1 Raspberry Pi & Camera Module

This system was selected due to its cost efficiency, flexibility for customization, ease of remote management, automation potential, and durability, which were superior to alternative devices such as smartphones.

3.2. Recording Method

3.2.1. Target Group

The participants were students of Laboratory of Social Systems Science, the Department of Environmental Systems, College of Systems Engineering and Science, Shibaura Institute of Technology. Students was required on Mondays and Fridays, and at least once from Tuesday to Thursday. They were allowed to choose their own working hours and seats, as the lab used a free-address system.

3.2.2. Recording Plan

The recording was conducted in a dining space located at the center of the Laboratory of Social Systems Science (Fig. 2). This space was designed to promote conversation rather than work and was frequently used for informal interactions, making it appropriate for analyzing everyday conversation.

Fig. 2 Dining area centered around the laboratory

Two Raspberry Pi units were mounted so that the faces of seated individuals could be captured (Fig. 3). These devices were secured near the ceiling by passing wire through insulation holes on the camera boards and adjusting the angles accordingly (Fig. 4).

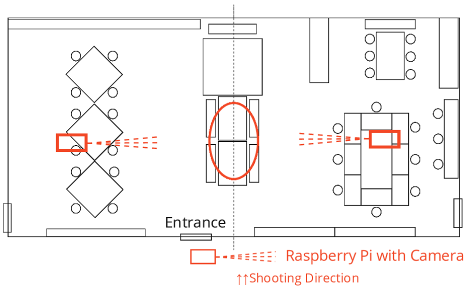

Fig. 3 Dining area adjacent to the laboratory

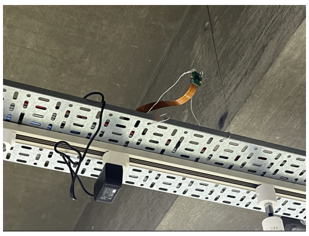

Fig. 4 Placement of Raspberry Pi camera

Recording was performed from 9:00 to 20:00, Monday through Friday, with a new video file saved every hour. The initial experiment took place between December 16 and 20, 2024, while the additional experiment was conducted between April 7 and 11, 2025.

3.2.3. Recording and Transmission Scripts

Shell scripts on the Raspberry Pi were used to manage both recording and transmission.

The video recording script used the rpicam command to capture 1080p video at 30 fps. Files were saved in MP4 format with filenames based on the recording date and time to ensure uniqueness. This process was automated using cron to run from 9:00 to 20:00.

Transmission to the analysis PC was implemented using the scp command over SSH. To allow passwordless operation, public key authentication was employed. File transfer was scheduled via cron to occur automatically at 20:00 each day.

The system described above is illustrated in (Fig. 5).

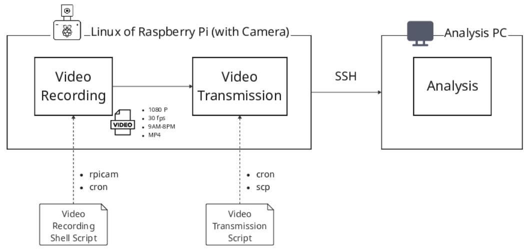

Fig. 5 System diagram

4. IDENTIFICATIONN

To correlate the analysis results with the productivity of each student, it was necessary to identify individuals in the video data. Face identification was performed by comparing detected faces with pre-collected facial photographs using deep learning techniques.

Photos were collected in a predefined format: five frontal images with neutral and smiling expressions, four oblique-angle images (top, bottom, left, right), and two profile images from 90-degree angles. For participants wearing glasses, an additional 2–3 photos were taken without glasses. In the initial experiment, 33 students participated, of whom 29 had usable photo data. In the additional experiment, 8 new students were added, bringing the total to 41, with 37 usable photo sets (Table 1).

Table 1 Conditions for Captured Photos

Photo Type	Number of Photos	Notes
Frontal photos	5	Various facial expressions such as neutral and smiling faces.
Diagonal photos	4	From oblique angles: upper, lower, left, and right.
Profile photos	2	Taken from 90 degrees to the left and right.
*For students wearing glasses, an additional 2–3 photos under the same conditions were taken without glasses.		

Each image folder was organized by individual. When a face was detected in the video, the most similar image in the pre-collected dataset was used for identification.

5. ANALYSIS_FLOW

The analysis was conducted using Python (version 3.10.2), running on Ubuntu 22.04 via WSL. GPU processing was used to speed up image processing.

5.1. Analysis Flow

5.1.1. Initial Experiment

Analysis flow for the initial experiment is shown in (Fig. 6).

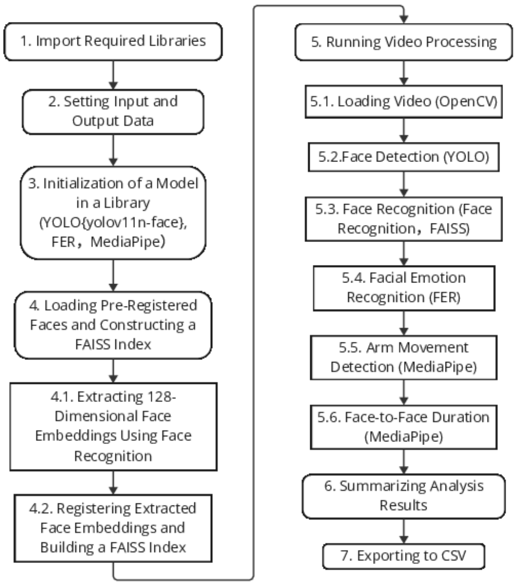

Fig. 6 Analysis flowchart for the initial experiment

First, the required Python libraries were imported and the input/output data settings were defined. Pre-trained models (YOLO, FER, MediaPipe) were initialized, and the pre-collected facial photos were loaded and converted into feature vectors. The YOLO model used for face detection (yolov11n-face) was specialized for that purpose. Afterward, video processing was executed, and each detected face was associated with its emotional state, arm movement, and face orientation. The results were aggregated and exported as CSV files.

5.1.2. Additional Experiment

The analysis flow for the additional experiment is shown in (Fig. 7).

The basic procedure was the same as in the initial experiment. However, an additional YOLO model (yolov8n) was employed to detect the number of people in the scene, which was then used to distinguish between working and communication contexts. Furthermore, the logic for detecting face orientation was updated. On the other hand, the arm movement detection used in the initial experiment was omitted, as it was considered potentially influenced by eating behavior in the dining space.

5.2. Library Configuration

5.2.1. Initial Experiment

In the initial experiment, video frames were extracted using OpenCV (4.10.0.84), and face detection was performed using YOLO (8.3.52). Face features were then computed using FaceRecognition (1.3.0) and FAISS (1.7.2) to identify individuals by comparing them to the pre-collected facial photos. Emotion analysis was performed using FER (22.5.1), while arm movements and the number of mutual face orientations were calculated

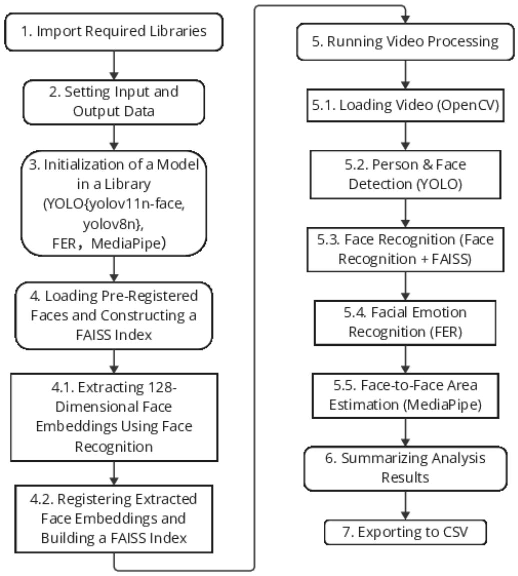

Fig. 7 Analysis flowchart for the additional experiment

using MediaPipe (0.10.20). These results were linked to the identified individuals and output in CSV format.

5.2.2. Additional Experiment

In the additional experiment, the logic for determining whether two people were facing each other was modified. First, eye landmarks were extracted using MediaPipe (0.10.20). Using these landmarks, gaze direction vectors were generated with the LineString class in Shapely (1.8.5). These vectors were compared with rectangular areas representing desk sections—divided into six vertical segments using the box class—to determine whether a person was facing someone. Additionally, the distinction between ”In Progress” and ”In Communication” was made based on the number of people detected using the YOLO person-detection model: one person indicated working alone, while two or more indicated communication.

5.2.3. Thresholds and Analysis Intervals

Thresholds and frame sampling intervals were adjustable in the analysis code. For the initial experiment, the face detection threshold was set to 0.6 (i.e., 60% confidence), and frames were sampled twice per second. Analysis was also performed using thresholds of 0.5 and 0.7, and an interval of 10 samples per second, resulting in a total of six combinations. In the additional experiment, the person detection threshold was set to 0.8, and the sampling interval was 10 times per second. The face detection threshold was lowered to 0.3, as using 0.7 caused frequent detection failures.

5.3. Output Results

The output results are as follows (Table 2).

The output includes, for each individual, the ratio of each detected emotion in the video (happy_ratio, surprise_ratio, angry_ratio, disgust_ratio, fear_ratio,

Table 2 Item Names and Supplementary Rules

Item Name	Supplementary Rules
happy_ratio	Appearance rate of each emotion
surprise_ratio	
angry_ratio	
disgust_ratio	
fear_ratio	
sad_ratio	
neutral_ratio	
p_to_n	"Number of emotional transitions when classified as Positive and Negative"
n_to_p	
p_to_p	
n_to_n	
arm_movements	Number of arm movements
face_time	Time spent facing each other
facing_area	Whether facing someone
In_Progress	Only one person captured
In_Communication	Two or more people captured
*Positive (happy, surprise), Negative (angry, disgust, fear, sad) ex) p_to_n=Positive→Negative	

sad_ratio, neutral_ratio), and the number of emotional transitions based on grouped emotions: Positive (happy, surprise) and Negative (angry, disgust, fear, sad). It also records the number of arm movements (arm_movement) and instances of mutual face orientation (face_time). Emotions are determined by the FER model, which analyzes facial landmarks and pixel features such as the eyes, mouth, and eyebrows to assign the most likely emotion.

In the additional experiment, the output also includes whether someone was present in the direction faced (facing_area), and whether the person was working (In_Progress) or communicating (In_Communication). This allows analysis to focus only on scenes involving communication.

All outputs are aggregated per individual.

5.4. Clustering of Results

K-means clustering was applied based on the number of times each student appeared in the video. Three clusters were generated, with Cluster 1 representing the group least frequently detected. The correlation between these clusters and weekly productivity was examined. Two productivity indicators were defined: the research achievement rate (completion of research-related tasks) and the overall task achievement rate (including non-research tasks), calculated by dividing the number of completed tasks by the number of assigned tasks.

6. RESULTS

The numerical results presented below are rounded to three decimal places.

6.1. Initial Experiment

The following results are based on communication data collected between December 16 and 20, 2024. The thresholds described here refer to face detection thresholds.

6.1.1. Threshold: 0.6, Analysis Interval: 2 per second

The analysis results for this configuration are shown in (Table 3, Table 4).

Table 3 Threshold 0.6, Analysis Interval: 2 per second — Correlation with Research Achievement Rate

Item	Cluster_1	Cluster_2	Cluster_3
arm_movements	0.05	-0.19	0.35
face_time	-	-	-
happy_ratio	-0.12	-0.45	-0.93
surprise_ratio	0.27	0.51	0.51
angry_ratio	0.22	0.31	-0.96
disgust_ratio	0.19	0.19	-0.04
fear_ratio	0.36	-0.59	-0.75
sad_ratio	0.39	0.034	-0.90
p_to_n	0.43	-0.54	0.12
n_to_p	0.43	-0.50	0.23
p_to_p	0.38	-0.40	0.28
n_to_n	0.045	-0.30	0.53

Table 4 Threshold 0.6, Analysis Interval: 2 per second — Correlation with Task Achievement Rate

Item	Cluster_1	Cluster_2	Cluster_3
arm_movements	0.40	-0.13	0.76
face_time	-	-	-
happy_ratio	0.22	-0.50	-0.98
surprise_ratio	0.17	0.45	0.69
angry_ratio	0.17	0.28	-0.94
disgust_ratio	0.15	0.16	-0.03
fear_ratio	0.33	-0.59	-0.76
sad_ratio	-0.31	-0.25	-0.89
p_to_n	0.14	-0.54	0.10
n_to_p	0.40	-0.50	0.38
p_to_p	0.38	-0.39	0.51
n_to_n	0.44	-0.26	0.31

Due to the camera angle, participants sitting close to the camera were not fully captured, making it impossible to analyze whether people were facing each other. As a result, all "face_time" outputs were 0, and correlations could not be computed.

No strong correlations (absolute value $\geq$ 0.50) were observed in Cluster 1. However, Clusters 2 and 3 exhibited several such correlations. In Cluster 2, the research achievement rate had strong correlations with surprise_ratio (0.51), fear_ratio (0.59), and p_to_n (0.53). In Cluster 3, the research achievement rate positively correlated with surprise_ratio (0.51), sad_ratio (0.77), and n_to_n (0.53), while showing negative correlations with happy_ratio (-0.93), angry_ratio (-0.95), and fear_ratio (-0.77).

For task achievement rate, Cluster 2 showed strong correlations with happy_ratio (-0.50), fear_ratio (0.56), p_to_n (-0.53), and n_to_p (-0.50). Cluster 3 had positive correlations with arm_movements (0.76), surprise_ratio (0.69), sad_ratio (0.82), and p_to_p (0.51), and negative correlations with happy_ratio (-0.98), angry_ratio (-0.91), and fear_ratio (-0.64).

These results suggest that surprise had a positive correlation and happy had a negative correlation with productivity in Clusters 2 and 3.

6.1.2. Other Thresholds and Intervals

By changing thresholds and analysis intervals while excluding "face_time", similar comparisons were conducted. When the threshold was set to 0.7 and the analysis interval to 10 per second (Table.5, Table.6) , the following findings were obtained:

Table 5 Threshold 0.7, Analysis Interval: 10 per second —Correlation with Research Achievement Rate

Item	Cluster_1	Cluster_2	Cluster_3
happy_ratio	-0.11	-0.78	-0.69
surprise_ratio	0.58	0.16	0.25
angry_ratio	0.26	0.43	0.022
disgust_ratio	0.27	0.41	0.17
fear_ratio	0.24	-0.13	0.17
sad_ratio	0.04	0.42	-0.28
p_to_n	-0.17	0.34	0.56
n_to_p	-0.18	0.34	0.55
p_to_p	-0.32	0.12	0.42
n_to_n	-0.15	-0.06	0.70

Table 6 Threshold 0.7, Analysis Interval: 10 per second — Correlation with Task Achievement Rate

Item	Cluster_1	Cluster_2	Cluster_3
happy_ratio	0.35	-0.81	-0.70
surprise_ratio	0.16	-0.09	0.18
angry_ratio	0.01	0.13	-0.07
disgust_ratio	0.04	0.59	0.36
fear_ratio	0.17	-0.24	-0.08
sad_ratio	-0.35	0.38	-0.30
p_to_n	0.40	0.27	0.40
n_to_p	0.38	0.27	0.39
p_to_p	0.27	0.23	0.20
n_to_n	-0.04	-0.35	0.56

In Cluster 1, the only strong correlation was between surprise_ratio and research achievement_rate (0.58). In Cluster 2, happy_ratio showed a strong negative correlation with both research (-0.78) and task achievement (-0.81), while disgust_ratio (0.59) had a positive correlation with task achievement. Cluster 3 showed correlations between happy_ratio (-0.69), p_to
n (0.56), n_to_p (0.55), and n_to_n (0.70) with research achievement, and between happy_ratio (-0.70) and n_to_n (0.56) with task achievement.

Increasing the analysis interval from 2 to 10 per second led to a greater number of correlated indicators.

6.2. Additional Experiment

The following results were obtained from data between April 7 and 11, 2025. The face detection threshold was 0.3, the person detection threshold was 0.8, and the analysis interval was 10 per second (Table 7, Table 8) .

Table 7 Threshold (FaceID: 0.3, Person Detection: 0.8), Analysis Interval: 10 per second — Correlation with Research Achievement Rate

Item	Cluster_1	Cluster_2	Cluster_3
facing_area	-	-	-
angry_ratio	0.26	-0.31	-
disgust_ratio	-0.10	0.15	-
fear_ratio	-0.60	-0.01	-
happy_ratio	-0.17	0.33	-
sad_ratio	-0.31	-0.22	-
surprise_ratio	0.27	0.32	-
p_to_n	-0.03	-0.66	-
n_to_p	-0.02	-0.71	-
p_to_p	0.08	-0.88	-
n_to_n	-0.01	-0.97	-

Table 8 Threshold (FaceID: 0.3, Person Detection: 0.8), Analysis Interval: 10 per second — Correlation with Task Achievement Rate

Item	Cluster_1	Cluster_2	Cluster_3
facing_area	-	-	-
angry_ratio	0.03	0.93	-
disgust_ratio	0.08	0.67	-
fear_ratio	-0.33	0.78	-
happy_ratio	0.13	0.53	-
sad_ratio	-0.38	-0.62	-
surprise_ratio	0.04	0.53	-
p_to_n	0.23	1.00	-
n_to_p	0.24	0.99	-
p_to_p	0.37	0.92	-
n_to_n	0.05	0.41	-

"Facing_area" could not be analyzed due to failed detection of other individuals in the gaze direction, and correlations could not be calculated as in the initial experiment. Cluster 3 only had two individuals, making correlation analysis unfeasible.

In Cluster 1, fear_ratio had a negative correlation with research achievement (-0.60), but no indicators showed a correlation of ≥ 0.50 with task achievement. In Cluster 2, strong negative correlations with research achievement were observed in p_to_n (-0.66), n_to_p (-0.71), p_to_p (-0.88), and n_to_n (-0.97). Strong positive correlations with task achievement were found in angry_ratio (0.93), disgust_ratio (0.67), fear_ratio (0.78), happy_ratio (0.53), surprise_ratio (0.53), p_to_n (1.00), n_to_p (0.99), and p_to_p (0.92), while sad_ratio had a negative correlation (-0.62).

Compared to the initial experiment, emotional transitions showed stronger correlations with task achievement, while high correlations with emotions like surprise or happy—found in the initial experiment—were not observed.

7. DISCUSSION

This study examined how nonverbal communication in lab conversations relates to weekly productivity.

7.1. Initial Experiment

7.1.1. Cluster 2

Research achievement positively correlated with surprise, fear, and p_to_n. Surprise may reflect active engagement, while fear suggests focused listening. Task achievement, however, negatively correlated with happy expressions and emotional transitions, possibly due to perceived frivolity or distraction.

7.1.2. Cluster 3

Similar trends appeared, with surprise and sad positively correlating with performance—sad possibly reflecting concentration. Angry had negative effects, while n_to_n and arm_movements positively correlated with research and task outcomes, though the latter may include unrelated actions like eating.

7.1.3. General Trends

Nonverbal effects varied by cluster, with frequent lab attendees showing stronger correlations. Increased surprise and reduced happiness were generally linked to higher productivity.

7.1.4. Thresholds and Intervals

A 10 FPS analysis interval improved sensitivity. Thresholds of 0.5–0.7 showed little difference; 0.7 offered optimal balance.

7.2. Additional Experiment

Focusing only on communication scenes, no strong emotion correlations emerged—suggesting effects in the initial experiment were task-context dependent. Emotional transitions correlated negatively with research and positively with task outcomes, implying that stable emotions aid advice reception, while expressiveness aids collaboration.

Cluster differences in fear trends indicate attendance influences interpretation—familiarity may turn emotion into a facilitator rather than a barrier.

7.3. Comparison with Existing Studies

While prior work focuses on formal meetings, this study highlights the productivity relevance of casual, daily conversations. Non-contact video analysis enabled low-burden monitoring of multiple individuals, offering a scalable and practical approach.

8. CONCLUSION

This study quantitatively examined how nonverbal cues—such as facial expressions, gaze, face direction, and group size—relate to weekly task achievement in lab conversations. In the initial experiment, surprise correlated positively with research achievement, while happiness showed a slight negative trend, suggesting that serious, discovery-driven expressions may enhance outcomes.

The additional experiment enabled scene classification based on gaze and spatial layout, revealing that emotionally stable interactions supported research advice, whereas expressive conversations aided task progress.

However, this study only identified correlations, not causality. Future research should test causal effects under controlled conditions. Technical limitations include gaze estimation accuracy and face recognition errors. Potential improvements include adjusting camera positions, refining detection logic, increasing facial data, and tuning thresholds. Current clustering relies on detection frequency, but future models should account for individual traits and social roles. To apply this method in real settings, intuitive feedback systems and real-time behavior support tools must be developed to address the lack of daily communication analysis technologies.

REFERENCES

[1] ALL DIFFERENT Inc., "Learning Agency: Survey on the Impact of COVID-19," https://prtimes.jp/main/html/rd/p/000000052.000005749.html (2024-12-28).

[2] Nikkei BP Consulting, "Overcoming Communication Distance [Part 2]: Urgent Survey of 1500 People on How Corporate Communication Changed During COVID," https://consult.nikkeibp.co.jp/ccl/atcl/20200525_1/ (2024-12-28).

[3] Japan Management Association, "Report on Organizational Stress Management," https://noma.actibookone.com/content/detail?param=eyJjb250ZW50TnVtIjoxMjg1NTUsImNhdGVnb3J5TnVtIjoxMjEwNX0=pNo=1 (2024-12-28).

[4] NTT TechnoCross, "ForeSight Voice Mining," https://www.ntt-tx.co.jp/products/foresight_vm/ (2024-12-28).

[5] muRata INNOVATOR IN ELECTRONICS, "NAONA × Meeting: Communication Visualization for 1-on-1 Meetings," https://www.murata.com/ja-jp/news/other/other/2019/0617 (2024-12-28).

[6] T. Abumi, "Characteristics and Roles of Group Dialogue in Science Learning," Journal of Science Education, vol. 64, no. 2, pp. 113–124, 2023.

[7] K. Nonaka, H. Yamashita, H. Hotta, M. Goto, "Network Construction Method Based on Communication Data on Business Chat Applications," Journal of Japanese Society for Artificial Intelligence, vol. 37, no. 2, pp. E-L63 1–11, 2022.

[8] Y. Tomita, S. Suzuki, Y. Mitsukura, "Development and Application of an Earphone-Type Pulse Wave Sensor," The Institute of Electrical Engineers of Japan, vol. 137, no. 2, pp. 236–241, 2017.

[9] Raspberry Pi, https://www.raspberrypi.com/ (2024-12-28).