Biological and Behavioral Information-based Method of Predicting Listener Emotions toward Speaker Utterances during Group Discussion

Motoki Sakai, Masaki Shuzo, Masahide Yuasa, Kanae Matsui, and Eisaku Maeda

Abstract There are many types of learning environments presented in higher education venues, requiring the development of a diverse repertoire of learning abilities. Group discussion (GD) is one such learning environment, and the students who participate require multiple communication skills. Technically, it is desirable for a student participating in a GD to understand other participants' emotional reactions toward their utterances to improve their locution, content, etc. The purpose of this research is to predict listeners' emotions in response to speakers' utterances using multimodal sensors. In experiments, GDs were conducted with 20 students. Six basic emotions were recorded as responses to speakers' utterances during the GDs using an emotional annotation tool. This study predicted the occurrence of the emotions by using an accelerometer, an electrocardiogram (ECG), and an electromyography (EMG). From sensor data, 56 features in the time and frequency domains were calculated, and Kruskal-Wallis tests and multiple comparison tests were performed to investigate whether there were significant differences among the features collected. As a result, there were significant differences among the groups of six basic emotions (p<0.01). As an applications, it has been shown that negative and positive emotions could be distinguished by support vector machine (SVM) with 76% F1.

Motoki SAKAI
Tokyo Denki University, 5 Senju Asahi-cho, Adachi-ku, Tokyo, Japan. Zip code: 120-8551, Japan
e-mail: m-sakai@mail.dendai.ac.jp

Masaki Shuzo
Tokyo Denki University, 5 Senju Asahi-cho, Adachi-ku, Tokyo, Japan. Zip code: 120-8551, Japan

Masahide Yuasa
Shonan Institute of Technology, 1-1-25 Tsujidonishikaiga, Fujisawa, Kanagawa, 251-8511, Japan

Kanae Matsui
Tokyo Denki University, 5 Senju Asahi-cho, Adachi-ku, Tokyo, Japan. Zip code: 120-8551, Japan

Eisaku Maeda
Tokyo Denki University, 5 Senju Asahi-cho, Adachi-ku, Tokyo, Japan. Zip code: 120-8551, Japan

1 Introduction

In venues of higher education, many types of learning environments (e.g., classroom lectures, labs, tutorials, seminars, and group discussions (GD)) are employed by instructors. Recently, there have been many attempts at quantifying the contextual variables of student learning to predict comprehension and adaptation capabilities and to improve methods. Copious data have been obtained and analyzed from real or staged lectures using information and communication technologies and biofeedback devices. Dr. Konomi et al. assembled a massive database from actual lectures using an e-learning tool to provide an analytical basis for improving educational programs and streamlining learning activities [1, 2]. Their result was a learning support system that utilized big-data concepts to derive patterns and variables of speaker-learner dynamics. This led to the concept of learning analytics (LA). Dr. Ogata et al. developed a learning support system to assemble and analyze similar data both inside and outside the classroom environment, providing another LA tool for big data [3]. Indeed, it has been reported that LA systems were used to increase student retention and achievement [4]. Furthermore, LA has been adopted by several universities (e.g., University of Maryland, Edith Cowan University, and Nottingham Trent University) for this purpose [4].

For this study, it is important to note that a very important component of the collected data focuses on biological and behavioral variables. In the literature [5], the synchronization of activities between a presenter and an audience in a laboratory setting was investigated using accelerometers attached to the subjects. Dr. Schleifer et al. focused on electromyography (EMG) signals to collect mental-stress data during computer operations [6]. The literature [7] has also measured student stress during classroom lectures by analyzing heart-rate variability (HRV). Dr. Bligh assessed attention durations during lectures using HRV tools [8].

In the prior studies mentioned above, various types of learning environments were applied. In this paper, GDs are focused. GDs are adaptable to many situations (e.g., decision-making fora, brainstorming sessions, and job interviews). GDs require versatile communication skills from all participants. Thus, it is desirable that these communication skills be improved by applying effective training techniques.

To date, there are many studies that evaluated relationship between verbal/nonverbal activities and level of communication skill, or functional roles in GD. In particular, gestures, line of sight, and contents of utterances have been often used to evaluate communication skill or process of communication [9, 10, 11]. Recently, gestures or line of sight has been measured with an acceleration sensor or eye tracking device, and nonverbal activities have been quantitatively assessed [12, 13]. On the other hand, several researches have evaluated student's concentration, stress, and emotion during e-learning and class room lecture with some sensors such as skin impedance [14], thermometer [15], EMG [6], ECG [7], and eye-tracking device [16]. Besides, there are studies that evaluated emotions during daily conversation with bio-sensor [17]. These prior studies have evaluated the degrees of student emotions or stress using averaged features computed at regular intervals. However, there are few studies which evaluate emotions toward individual utterances in GD by using sensors. Thus, we focus on each utterance and reaction behavior of GD participants to obtain direct improvement opportunities related to communication skills. This study proposes a method of predicting listener emotions toward each speaker utterance during GD using multimodal sensors. The previous studies (e. g. [18]) presented that there is relationship between aggregability and productive group work including GD. To increase aggregability, it is important to understand group member's emotions. To fine tune corrective measures, however, it is important to capture and analyze individual participants' emotions in relation to specific cues. Moreover, these cues are precisely what should be used by speakers to gauge the receptibility of their utterances. Most people do this subconsciously to improve locution, amenability, levels of complexity, and more. However, most people lack finely tuned tools that can help them get it right and make potent tweaks to their delivery. In the following sections, we present such a system to generate annotations related to listener emotions in response to each speaker utterance during a GD. Furthermore, we demonstrate the feasibility of predicting said emotions using multimodal sensors, which should lead to effective adaptation tools.

2 Experiment

2.1 Group discussions

The GD experiments were conducted with the participation of 20 Japanese university and postgraduate subjects (16 male), aged 19-23, assembled from five higher-level institutions. Subjects were divided into five groups of four each. Three groups were tasked to perform a brainstorming session to answer, "What is an effective teaching technique for programing education at an elementary school?" (GD1). The remaining two groups were tasked (three times each) to conduct GDs aimed at building consensus (GD2). In summary, there were nine total GD sessions.

For GD1, the experimenter participated in the GD as a facilitator and explained the rules. In the GD2, the experimenter did not participate, but explained the rules. In each GD, the four participating students made desultory conversation for 5 min as an ice breaker prior to the actual start. Then, when the respective GD began, the experimenter provided the agenda. Each GD lasted 30 min. The experimenter of GD1 facilitated but did not directly participate. Only the four students participated in GD2. During each session, the students were instructed to address each other by "A," "B," "C," or "D" as assigned.

2.2 Adopted sensors

Sensors, cameras, and environmental measurement instruments were used for each GD, as shown in Table 1. The sensors recorded participants' vitals, verbal/nonverbal metrics, and environmental data. An electrocardiogram (ECG), an EMG (left and right trapezius: Fig. 1), and an electroencephalogram (EEG) were applied at a 1,000-Hz sampling rate with 16-bit resolution. The sampling rate of acceleration was 100 Hz. The ECG signal was measured using a modified CM5 lead, as shown in Fig. 2. A positive electrode was located on the 5th rib corresponding to V5 in the 12-lead ECG, and the negative was attached below the right clavicle. The grand electrode was attached to the lower abdomen. The EEG signals were measured from the four students via Fp1 and Fp2, according to the international 10/20 system. In the following feature extraction and evaluation steps, only acceleration attached on the chest, ECG, and EMG signals were adopted.

Table 1 Adopted sensors for obtaining biological and behavioral information

Sensor name	Measurement object	Location of sensor	Measurement purpose
TSND151	Acceleration, gyro	Both wrists, chest, top of head	Grasp of non-verbal information
AMP151	EEG、ECG, EMG	EEG:Fp1,Fp2、 ECG: modified CM5, EMG: trapezius	Grasp of automatic nervous activity
Fitbit Ionic	Pulse wave	Non-dominant arm	Grasp of automatic nervous activity
JINS MEME	EOG	Face	Grasp of vitality
Video Camera	Verbal/non-verbal information	Center of table, around table	Grasp of verbal/non-verbal information
Netatmo	CO2 concentration, humidity, temperature	Several places in experiment room	Grasp of environmental information

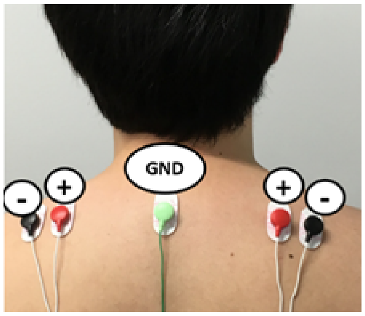

Fig. 1 Electrode positions of trapezius EMG

2.3 Generation of emotional annotations

This research defines emotional annotation as a listener's emotions expressed in response to a speaker's utterance. Six basic emotions were tracked as emotional

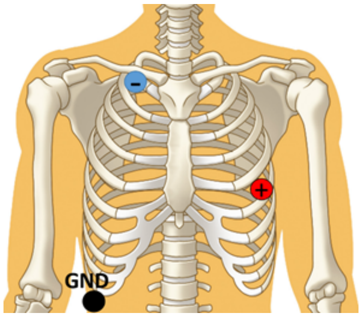

Fig. 2 Electrode position of ECG

labels [19]. To generate an emotional annotation, we used the tool shown in Fig. 3. During the experiments, video data were recorded via YouTube live streaming. During each GD, the experimenter manually generated an utterance record for every speaker's utterance as shown in Fig. 3 using the YouTube chase playback feature. For each utterance record shown in Fig. 3, '02:09. 13 65', '02:09. 20 65', or '02:09. 52 65' are time stamps corresponding to a speaker utterance. In the video-check area of Fig. 3, the video image was advanced at rate corresponding to the time stamp of the click of an utterance record. Thus, a selected speaker's utterance could be checked.

Promptly after each GD, the four participants created emotional annotation data for each utterance. Each student clicked upon each utterance record and recaptured the emotions felt during the GD session. Then, the recalled emotions were recorded using buttons corresponding to the six basic ones (i.e., joy, disgust, surprise, sadness, anger, and fear) (Fig. 3). Each emotion was scaled from 0 to 4, where "0" means negligible emotion felt, and "4" means strong emotional reaction. "0" was also suitable for "no emotion felt." This process was not applied for subjects' own utterances. In the factor area of Fig. 3, participants selected the factor (i.e., content, volume, mood, cadence, utterance rate, gestures, and others) that elicited an emotional reaction, per measure. In the note area of Fig. 3, participants were tasked to write their speaker alpha name. These operations were repeated for all GDs so that we could obtain the emotional annotation data as described. Data were recorded to a .csv file, and four files were generated for each GD session, corresponding to subjects.

Our research group conducted this experiment obtaining ethical approval from the ethical review committee of Tokyo Denki University. Before this experiments were conducted, we got subject's informed consent. Firstly, a purpose of this research and experiment procedure were stated. Next, we explained that movie data and obtained sensor data will be distributed to researchers who hope to use these data set. Finally, subjects signed a letter of consent if they could understand a purpose of this research and so on. In addition, we explained that subjects can drop a permission for the usage of their data set at a later date if they desire.

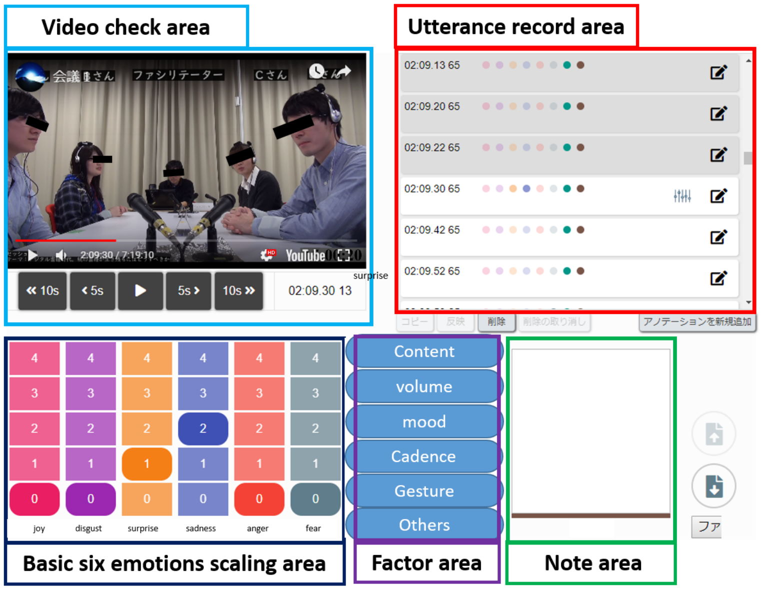

Fig. 3 Creation tool for the emotional annotations

3 Features of biological and behavioral information

There were 42 features captured from the acceleration signals [20, 21]: 10 ECG [22, 23], and four EMG [24]. 56 features total were adopted to classify the six basic emotions.

For the acceleration signal, the summation of vector norm (SVN); the signal-magnitude area (SMA); the root mean square (RMS) of the X, Y, and Z axes (RMS X, RMS Y, RMS Z); the curve length (CL) of the X, Y, and Z axes (CL X, CL Y, CL Z); the spectral summation (SS) of the X, Y, and Z axes (SS X, SS Y, SS Z); the maximum power (MP) in the spectral domain (MP X, MP Y, MPZ); the frequency representing maximum power (FMP) in the spectral domain (FMP X, FMP Y, FMP Z); the standard deviation (STD) of the acceleration signal of the X, Y, and Z axes (STD X, STD Y, STD Z); the mean of the acceleration signal of the X, Y, and Z axes (MEAN X, MEAN Y, MEAN Z); the median of the acceleration signal of the X, Y, and Z axes (MEDIAN X, MEDIAN Y, MEDIAN Z); the differential entropy of the vector norm in the time domain (ENTROPYT); the entropy of the vector norm in the frequency domain (ENTROPYF); the mean of the vector norm (VN) (MEAN VN); STD of VN (STD VN); the maximum value of VN (MAX VN); the minimum value of VN (MIN VN); the differential value between MAX VN and MIN VN (RANGE); the median of VN (MEDIAN VN); the summation of VN (SUM VN); the variance sum (VAR(X+Y+Z)); the absolute value three axes (ABS); the interquartile range

(IR); the kurtosis of the vector norm (KURTOSIS); the skewness of the vector norm (SKEWNESS); and the median crossings (MC) were used. The SVN, SMA, RMS X (RMS Y, RMS Z), CL X (CL Y, CL Z), ENTROPYT, ENTROPYF, VAR(X+Y+Z), ABS, IR, and MC were calculated using Eqs. (1), (2), (3), (4), (5), (6), (7), (8), (9), and (10), respectively.

$$SVN = \frac{1}{N}\sum_{i=1}^{N}\sqrt{x_i^2 + y_i^2 + z_i^2} \tag{1}$$

$$SMA = \frac{1}{T}\Big(\int_0^T |x_t|dt + \int_0^T |y_t|dt + \int_0^T |z_t|dt\Big) \tag{2}$$

$$RMSX = \sqrt{\frac{1}{N}\sum_{i=1}^{N}x_i^2} \tag{3}$$

$$CLX = \sum_{i=2}^{N}|x_{i-1} - x_i| \tag{4}$$

$$ENTROPYT = \int_X f(x)log\big(f(x)\big)dx \tag{5}$$

$$ENTROPYF = \sum_{i=1}^{N/2} p_i log_2(p_i) \tag{6}$$

$$Var(X{+}Y{+}Z) = Var(X) + Var(Y) + Var(Z) + 2Cov(X,Y) + 2Cov(X,Z) + 2Cov(Y,Z) \tag{7}$$

$$ABS = \frac{\sum_{i=1}^{N}|x| + \sum_{i=1}^{N}|y| + \sum_{i=1}^{N}|z|}{3N} \tag{8}$$

$$IR = percentile(75, VN) - percentile(25, VN) \tag{9}$$

$$\begin{cases} t_i = VN_i - median(VN) \\ MC = \sum_{i=1}^{N-1} sgn(t_i - t_{i+1}) \\ sgn(t_i, t_{i+1}) = \{1 : if(t_i * t_{i+1}) < 0,\ 0 : if(t_i * t_{i+1}) > 0\} \end{cases} \tag{10}$$

where N is the total number of samples; x_i, y_i, and z_i are the ith samples of the x, y, and z axes, respectively; X, Y, Z are the total signals of the x, y, and z axes, respectively; and VN is the vector norm of the acceleration signal. To compute SS X, SS Y, and SS Z, raw acceleration signals were transfoemd into frequency domain with the fast Fourier transform (FFT), and sum of all power in frequency domain was obtained. In a similar way, MP X, MP Y, MP Z, FMP X, FMP Y, and FMP Z were calculated after the FFT.

For the ECG signal, features of both time and frequency domains were calculated. Before calculating the features, raw ECG signals were filtered using a 0.5-150 Hz band-pass filter; the positions of R-waves were detected using the wavelet transformation modulus-maxima method [25]; and the heart-rate variability signals were generated. In the time domain, the mean peak value of detected R waves (ECG AMP), the mean of the heart rate (HR), and the RMS of the successive differences between contiguous R–R intervals (RMSSD) were used. In the frequency domain, summation of frequency powers whose frequency bands were lower than 0.4 Hz (TOTAL POWER); the summation of spectral powers in the very-low frequency (VLF) band (0.033-0.04 Hz); the summation of spectral powers in the low frequency (LF) band (0.04-0.15 Hz); the summation of spectral powers in the high frequency (HF) band (0.15-0.4 Hz); and the ratio between LF and HF (LF/HF), LF/TOTAL POWER - VLF (LF NORM), HF/TOTAL POWER - VLF (HF NORM) were adopted. These features reflect autonomic nerve activity, as shown in Table 2, and have often been used in studies of human stress [26]. These ECG features taken from the GDs were divided by ones at rest to diminish individual differences of heart activity.

Table 2 ECG features and correspondent autonomic activities

ECG features	Correspondent autonomic activities
RMSSD	Vagal tone
TOTAL POWER	Total activity of autonomic nerve
VLF	Slow activity of sympathetic nerve
LF	Activities of both sympathetic parasympathetic nervous
HF	Activity of parasympathetic nerve
LF/HF	Balance between parasympathetic and sympathetic nervous
LF NORM	Activity of sympathetic nerve
HF NORM	Activity of parasympathetic nerve

For the EMG signal, the summation of the absolute value of the EMG signal recorded on the right and left trapeziuses (EMG AMP (Right), EMG AMP (Left)), and the summation of EMG's spectral powers in the LF band (170-200 Hz) (EMG LF (Right), EMG HF (Left)) were adopted. Additionally, EMG features captured during GDs were divided by ones at rest to diminish individual differences of muscle strength. To compute ECG and EMG features in frequency domain, the FFT was adopted.

The above-described 56 features were computed for each recorded annotation. Concretely, 7s signals after 4 s from the start of recorded annotations were extracted to compute features. This time interval was determined arbitrarily and could be improved, because the length of utterances differ from each other. From the three GD1 and six GD2 sessions, 934 joy, 76 disgust, 434 surprise, 64 sadness, 29 anger, and 158 fear annotations were obtained. 56 features were calculated for a total of 1,695 annotations. In this paper, these annotations labels were counted when the scaling value was > 0. Furthermore, magnitude of scaling values (0-4) were not considered in the following evaluation step.

4 Statistics Analysis for 56 features

In this paper, the Kruskal-Wallis test was used to assess whether there were statistically significant differences among computed features associated with the six basic emotions. In general, a normality is tested at first, and a method to test differences among groups is determined. In this study, we have 56 features, and it can be supposed that some are normally distributed and others are non-normally distributed (See Figs. 5 through 9 shown in below). Therefore, we adopted the nonparametric Kruskal-Wallis test to test differences among six groups, because this test can be used for normally and non-normally distributed data.The results are shown in Table 3. In 36 of the 56 features, there were statistically significant differences among the six basic emotions (p<0.01).

Next, the Bonferoni method-based multiple comparison tests were performed for those 40 features. The feature yielding the most significant differences among the six basic emotions was selected for each acceleration, EMG, and ECG. The results of multiple comparison tests are shown in Tables 4-6

Including the above features, SUM VN, LF, and AMP LP (Left) results from multiple comparison tests for the 36 features were significantly different in the Kruskal–Wallis tests. Thus, they were rounded up. As a consequence of multiple comparison tests, the numbers of features showing significant differences of combination of two emotions (p<0.01) are shown in Table 7.

5 Application

As described in the introduction, the final purpose of this research is to construct a system to provide information related to listeners' emotions toward speaker's utterances in order to improve her/his locution. In this section, we present one example of machine learning application.

In previous section, the six basic emotions were evaluated, but the number of data set is not sufficient to classify the six basic emotions (For example, the number of disgust data set was76, and that of fear data set was 29.). Thus, we propose a simple method to discriminate negative emotion from positive one. In this study, joy is restated as positive emotion, remaining five emotions are regarded as negative emotions according to literature [19]. To perform a machine learning, we used GD2's data for learning, and used GD1's one for test. In particular, 338 negative emotion and 507 positive emotion data were used as learning data set, and 147 negative emotion and 93 positive emotions data were used as test data set.

Table 3 *p*-values of the Kruskal-Wallis test among the six basic emotions with 56 features

SVN	SMA	RMS X	RMS Y	RMS Z	CL X	CL Y
0.000	0.083	0.000	0.364	0.000	0.000	0.000
CL Z	SS X	SS Y	SS Z	MP X	MP Y	MP Z
0.000	0.000	0.000	0.000	0.000	0.445	0.000
FMP X	FMP Y	FMP Z	STD X	STD Y	STD Z	MEAN X
0.240	0.629	0.177	0.000	0.000	0.001	0.000
MEAN Y	MEAN Z	MEDIAN X	MEDIAN Y	MEDIAN Z	ENTROPYT	ENTROPYF
0.000	0.000	0.000	0.000	0.000	0.000	0.000
MEAN VN	STD VN	MAX VN	MIN VN	RANGE VN	MEDIAN VN	SUM VN
0.000	0.000	0.056	0.000	0.000	0.000	0.000
VAR(X+Y+Z)	ABS	IR	KURTOSIS	SKEWNESS	AUTOCORR	MC
0.000	0.083	0.006	0.037	0.819	0.000	0.000
TOTAL POWER	VLF	LF	HF	LF NORM	HF NORM	LF/HF
0.053	0.034	0.000	0.002	0.000	0.414	0.005
ECG AMP	HR	RMSSD	EMG AMP (Left)	EMG AMP (Right)	EMG LF (Left)	EMG LF (Right)
0.599	0.073	0.063	0.001	0.043	0.000	0.001

As a machine learning algorithm, support vector machine (SVM) was adopted in this paper, and the linear function was selected as a kernel function because this

Table 4 Result of multiple comparison test for SUM VN

Comparison		Mean of difference	95% Confidence interval		*p*
			Lower	Upper	
Joy	Disgust	356.26	185.24	527.30	0.000
Joy	Surprise	188.59	96.42	280.76	0.000
Joy	Sadness	622.81	439.59	806.03	0.000
Joy	Anger	530.83	203.99	857.67	0.000
Joy	Fear	193.57	58.75	328.38	0.000
Disgust	Surprise	-167.67	-347.06	11.73	0.091
Disgust	Sadness	266.55	27.27	505.83	0.016
Disgust	Anger	174.57	-186.69	535.83	1.000
Disgust	Fear	-162.69	-367.29	41.91	0.294
Surprise	Sadness	434.22	243.16	625.27	0.000
Surprise	Anger	342.24	10.940	673.53	0.034
Surprise	Fear	4.9762	-140.31	150.27	1.000
Sadness	Anger	-91.98	-459.17	275.22	1.000
Sadness	Fear	-429.24	-644.14	-214.30	0.000
Anger	Fear	-337.26	-682.90	8.34	0.062

Table 5 Result of multiple comparison test for EMG LF (Left)

Comparison		Mean of difference	95% Confidence interval		*p*
			Lower	Upper	
Joy	Disgust	-121.70	-286.79	43.40	0.287
Joy	Surprise	-7.35	-87.74	73.05	0.999
Joy	Sadness	-116.02	-294.86	62.81	0.434
Joy	Anger	-282.98	-643.20	77.25	0.220
Joy	Fear	-144.92	-263.97	-25.87	0.007
Disgust	Surprise	114.35	-57.76	286.46	0.406
Disgust	Sadness	5.67	-229.15	240.49	1.000
Disgust	Anger	-161.28	-552.33	229.77	0.849
Disgust	Fear	-23.23	-216.44	169.99	0.999
Surprise	Sadness	-108.68	-294.01	76.65	0.551
Surprise	Anger	-275.63	-639.13	87.86	0.256
Surprise	Fear	-137.57	-266.18	-8.97	0.028
Sadness	Anger	-166.96	-564.00	230.09	0.838
Sadness	Fear	-28.90	-233.98	176.18	0.999
Anger	Fear	138.06	-235.89	512.01	0.900

Table 6 Result of multiple comparison test for LF

Comparison		Mean of difference	95% Confidence interval		*p*
			Lower	Upper	
Joy	Disgust	32.66	-95.47	160.79	0.979
Joy	Surprise	-101.23	-166.63	-35.83	0.000
Joy	Sadness	62.11	-87.69	211.91	0.846
Joy	Anger	-51.57	-330.87	227.74	0.995
Joy	Fear	-211.84	-310.79	-112.89	0.000
Disgust	Surprise	-133.89	-267.96	0.18	0.051
Disgust	Sadness	29.45	-160.64	219.54	0.998
Disgust	Anger	-84.23	-387.06	218.60	0.969
Disgust	Fear	-244.50	-397.76	-91.24	0.000
Surprise	Sadness	163.34	8.43	318.25	0.032
Surprise	Anger	49.66	-232.42	331.74	0.996
Surprise	Fear	-110.61	-217.14	-4.08	0.036
Sadness	Anger	-113.68	-426.29	198.94	0.906
Sadness	Fear	-273.95	-445.74	-102.16	0.000
Anger	Fear	-160.27	-451.96	131.42	0.621

Table 7 Number of features being capable of classify each combination of two emotions

Combination of emotions	Joy-Disgust	Joy-Surprise	Joy-Sadness	Joy-Anger	Joy-Fear	Disgust-Surprise	Disgust-Sadness	Disgust-Anger
Number of features	15	15	21	8	20	3	8	3

Combination of emotions	Disgust-Fear	Surprise-Sadness	Surprise-Anger	Surprise-Fear	Sadness-Anger	Sadness-Fear	Anger-Fear
Number of features	10	13	5	6	1	13	6

kernel yielded best result compared to other functions such as the Gaussian and polynomial kernels.

To distinguish negative and positive emotions, 56 features computed from acceleration, ECG, and EMG signals were used. To select effective feature set, the sequential feature selection (SFS) method [27] was used. SFS is performed according to the following four steps.

- I. SFS starts with an evaluation of the F1 value (mentioned below) of the estimation for each single feature, selecting the best one.
- II. SFS is performed using two features at a time, the one that maximized the F1 value in the first step and each of the other features from the remaining subsets, selecting the pair with the highest F1 value.
- III. SFS evaluates three features subsets, the two that were determined in the previous two steps, and another selected from the remaining subsets.
- IV. The procedures described in steps I–III are repeated.

In this research, we focus on detecting listeners' negative emotions toward speaker's utterances, because a speaker would be able to improve her/his inappropriate expression by finding listeners' negative emotions toward own utterances. Therefore, the case that negative emotion is detected as negative one is regarded as true positive (TP). In the same way, the case that positive emotion is detected as positive one is regarded as true negative (TN), the case that negative emotion is detected as positive one is regarded as false negative (FN), and the case that positive emotion is detected as negative one is regarded as false positive (FP). With these values, accuracy = (TP + TN) / (TP + FP + TN + FN), precision = TP / (TP + FP), recall = TP / (TP + FN), and F1 value = 2recall*precision / (recall + precision) are evaluated.

The evaluation result is shown in Fig. 4. Fig. 4 shows a relationship between four evaluation scores and the number of selected features. As Fig. 4, accuracy, precison, and F1 value reached maximum values (accuracy = 71%, precision = 77%, recall = 76%, and F1 value = 76%), when the number of selected features was five. Top five features SFS method selected were EMG AMP (Right), HF, LF, RMS Y, and IR in order of selection. Top five features' histograms are shown in Figs. 5, 6, 7, 8, and 9.

6 Discussion

Of the 56 features computed from acceleration, ECG, and EMG signals, the SUM VN obtained from the acceleration signal showed the most significant differences among the six basic emotions in the Kruskal–Wallis test (Table 3). Table 4 shows that joy was significantly different from the other five emotions, and shows an identifiability for surprise-sadness and sadness-fear. Table 5 shows that the EMG LF (Left) was less effective than the SUM VN, because only the combination of joy-fear was significantly different in the multiple comparison test. In the same manner, LF was less effective than SUM VN, as shown in Table 6. From these results, we

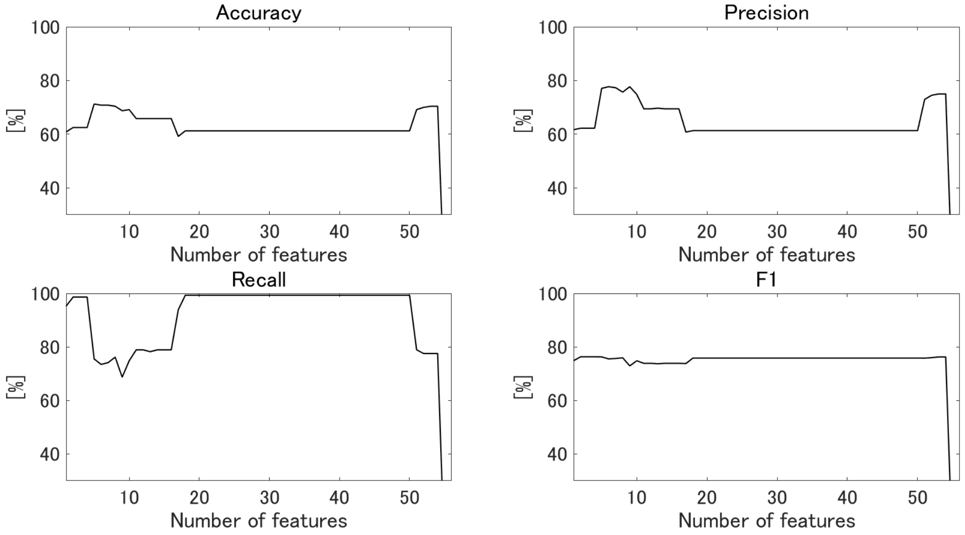

Fig. 4 Classification results

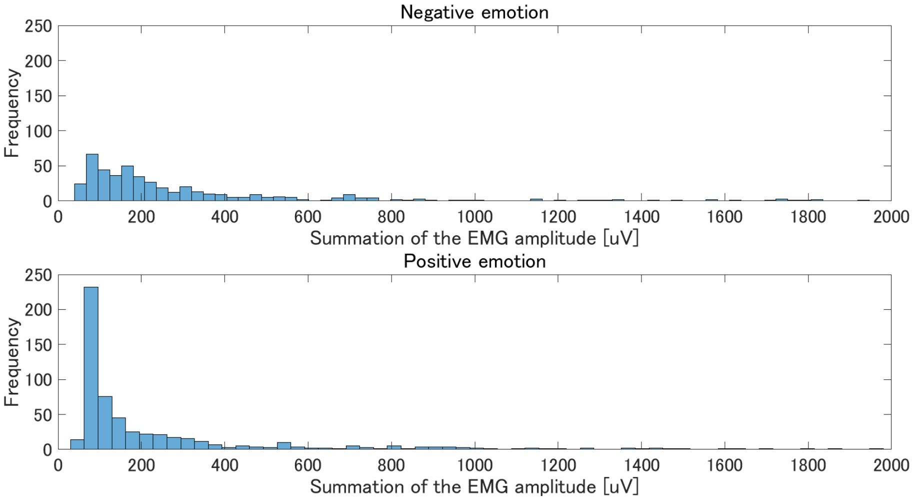

Fig. 5 Histograms of EMG AMP (Right), top: negative emotion group, bottom: positive emotion group

can surmise that one feature was limited for classification. However, it might be made feasible with the addition of other features. Furthermore, as a consequence of multiple comparison tests on the number of features, all combinations of the six basic emotions showed significant differences, as shown in Table 7. It also shows

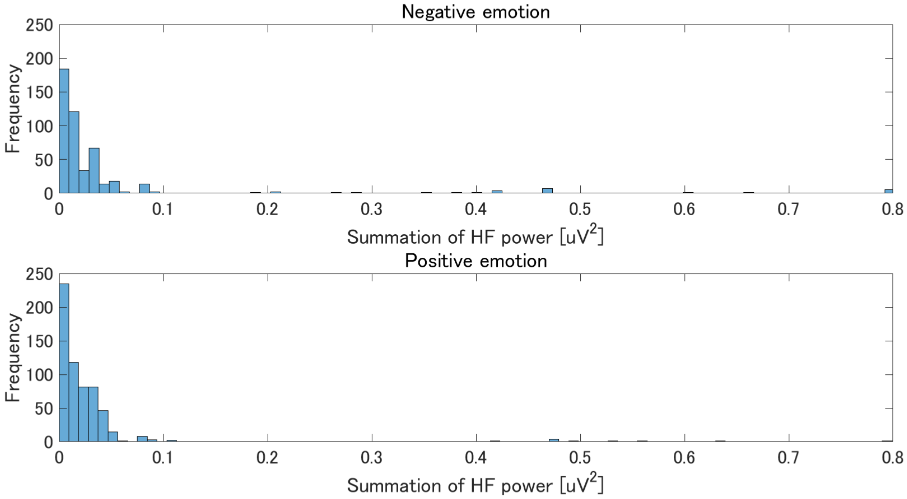

Fig. 6 Histograms of HF, top: negative emotion group, bottom: positive emotion group

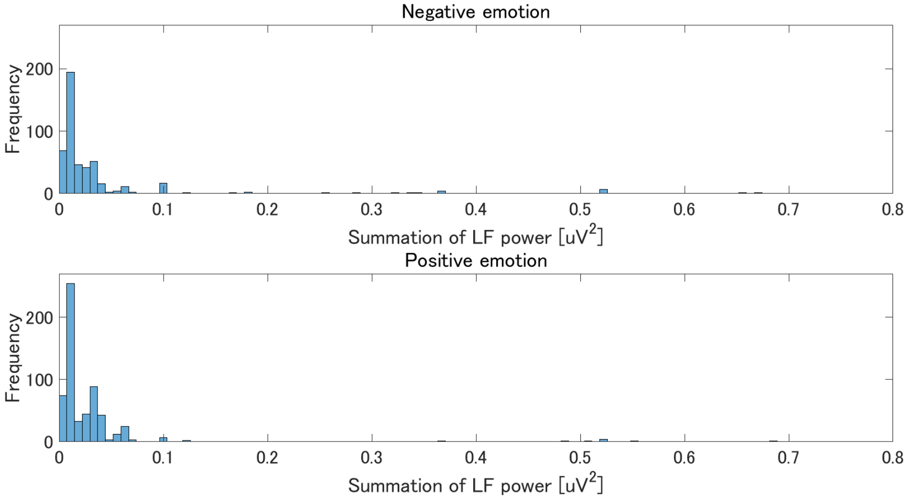

Fig. 7 Histograms of LF, top: negative emotion group, bottom: positive emotion group

that, in 15-21 features, significant differences existed between joy and disgust, joy and surprise, joy and sadness, and joy and anger were found. This also indicates that positive emotions might be able to be discriminated from negative ones. Although joy is a positive emotion and anger is negative, significant differences between them were found in only eight features, implying that it might be more difficult than assumed to find distinguish joy and anger. In many cases, anger was not significantly

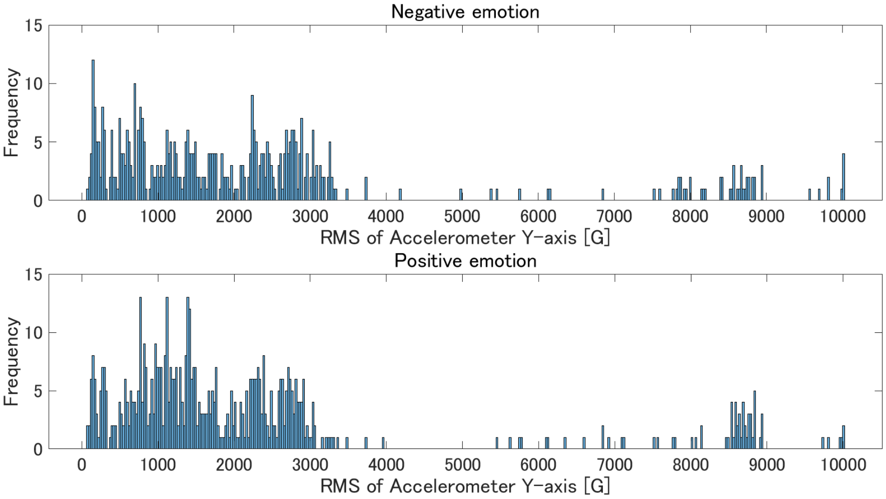

Fig. 8 Histograms of RMS Y, top: negative emotion group, bottom: positive emotion group

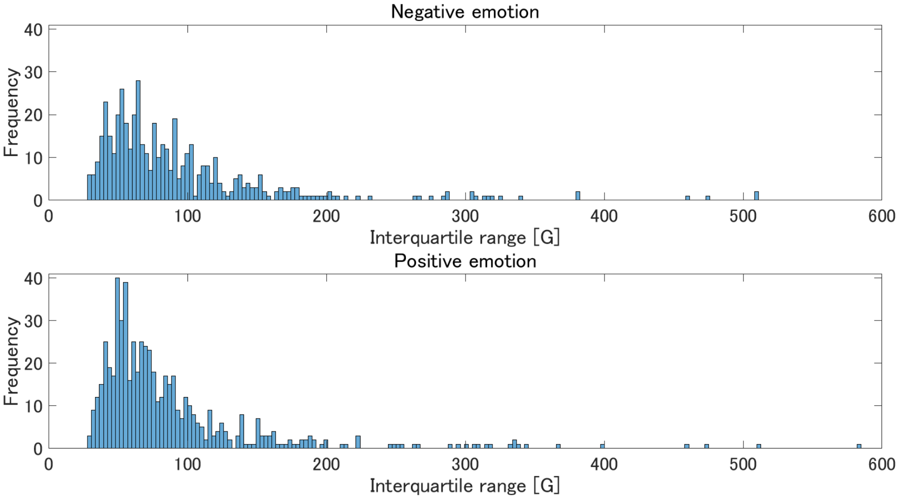

Fig. 9 Histograms of IR, top: negative emotion group, bottom: positive emotion group

different from other emotions, as shown in Table. 7. These results might have been caused by the number of emotional annotations of anger, where there were only 29 anger labels. Thus, this research could not accurately analyze the identifiability of anger. In this experiment, a brainstorming and consensus GDs were conducted. It is generally considered that students participating in GDs rarely express anger, especially during brainstorming. Therefore, new annotation labels more favorable to GDs may be needed for future experiments.

As for the SVM-based classification results, as described above, F1 value reached maximum when five features were selected. This result shows that the balance between precision and recall was optimized with five selected features. On the other hand, the recall reached maximum value 98% when the two features: EMG AMP (Right) and HF were selected. This result is also effective for the aim of this research even if precision obtained with two features was not sufficient (= 62%). To improve inappropriate expression according to predicted listener's emotion, not all of negative emotions have to be detected. Granted that several utterances evoking listener's negative emotion are not detected, utterances evoking positive emotion must not be recognized to be one to improve. In that sense, the SVM learning machine with EMG AMP (Right) and HF can be functional.

Next, we discuss what top five features selected by SFS method mean. In Fig. 5, distributions of EMG AMP (Right) were different between the negative and positive emotions groups. In the positive emotion group, the peak was located around 80 uV. On the other hand, summations of EMG amplitude were widely distributed throughout larger value. In other words, EMG AMP (Right) in negative emotion group took larger value compared to those in positive emotion group. In general, the amplitude of EMG increases under stress. In sum, it can be considered that listeners had negative emotion toward speaker's utterance and felt stress. The reason why only EMG AMP (Right) was selected instead of EMG AMP (Left) is unknown to us, but we surmise that this could be related to better arm. With respect to HF and LF, distributions between negative and positive emotions groups were similar to each other (Figs. 6 and 7). However, it is known that balance between HF and LF are related to activity of automatic nerve, and it can be supposed that HF and LF can predict listener's emotion through activity of automatic nerve. As with EMG AMP (Right), it is noted that HF and LF change under the influence of mental stress, and these selections were valid from physiological point of view. With respect to RMS Y, values of RMS Y have wide distribution both in positive and negative emotions groups, but distributions between positive and negative emotions groups were clearly different (Fig. 8). In positive emotion group, the peaks were located around 1000 G. On the other hand, there were two sets of peaks around 200-1000 and 2000-3000 G. We can assume that listeners having positive emotion expressed their emotion by moderate activity. The reason why RMS Y was selected from three axes is that activity in horizontal direction was not restricted compared to activities in vertical (X-axis) and anteroposterior (Z-axis) directions, because participants in GD sat in front of table. As to IR, the peak was located around 40 G in the positive emotion group, and the peak of negative emotion group was located around 60 G. (Fig. 9) This result means that listeners did not express their emotion by furious movement, which agrees with the result of RMS Y.

In the evaluation result of the Kruskal-Wallis test, it was indicated that features computed from accelerometer such as SVN and SMA would be promising to predict the six basic emotions. As it is, the biological signals such as EMG AMP (Right), HF, and LF were more contributive than other features computed from accelerometer in application based on the SFS and SVM. However, we cannot conclude that these results were contradictory. Assuredly, EMG AMP (Right), HF, and LF were more contributive to distinguish negative emotions from positive emotions. However, features computed from accelerometer including SVN would be more contributive to tell the six basic emotions. We are planning to assemble more data set for a prediction of the six basic emotions.

In this research, many sensors were attached to participants in GD. Given actual learning situation, ease of attachment of sensors, and motion artifacts, minimal sensor set should be selected. In this paper, an effectiveness of HF and LF was indicated. Therefore, we will use Fitbit to obtain features of HRV. Absolutely, the HRV signal obtained by the Fitbit can be less accurate than that computed by the ECG signal. In a future work, we will propose a smart watch-based method to predict negative and positive emotions in view of capability of a smart watch.

Although there were a few problems related to annotations such as anger, it can be concluded that the evaluation results reasonably indicate the possibility of classifying the six basic emotions with effective features derived from acceleration, ECG, and EMG signals. Especially, we indicated the possibility that two emotions: positive and negative can be distinguished with the SVM. This SVM-based-application can be effective to support participants in GDs by using prediction information.

7 Conclusion

The goal of this research was to predict listeners' emotional reactions toward speakers' utterances using multimodal sensors. GD brainstorming (GD1) experiments and GD consensus-building (GD2) experiments were conducted with 20 student participants. In this research, the six basic emotions were adopted as emotional labels, and listeners' emotions toward speaker utterances during GD1 and GD2 sessions were manually annotated using an emotional annotation tool. This research attempted to classify the six basic emotions during GD1 and GD2 by using acceleration, ECG, and EMG metrics. From these data, 56 features were computed, and the Kruskal–Wallis tests with multiple comparison tests were performed to investigate whether there were significant differences among the features within the six emotion groups, there were significant differences among features grouped into the six basic emotions, and the possibility to predict those emotions using the obtained signals was supported. As an example of application, negative and positive emotions were distinguished by SVM, and an effectiveness of features calculated from biological signals including EMG AMP (Right), HF, and LF were shown.

Acknowledgements This work was supported by the Grant-in-Aid for Cutting-Edge Research for the social implementation of next-generation artificial intelligence technologies of the New Energy and Industrial Technology Development Organization (NEDO) of Japan and Sciences Research Grant, Grant-in-Aid for Scientific Research (B) (General). We would like to thank Editage (www.editage.jp) for their English language editing services.

References

1. H. Ogata, S. Liu, K. Mouri: Ubiquitous Learning Analytics Using Learning Logs. Workshop Proc. LAK2014, (2014).
2. A. Shimada, Y. Taniguchi, F. Okubo, S. Konomi, H. Ogata: Online Change Detection for Monitoring Individual Student Behavior via Clickstream Data on e-Book System. Proc. LAK2018, pp. 446–450, (2018).
3. P. Ocheja, B. Flanagan, H. Ogata: Connecting Decentralized Learning Records: A Blockchain Based Learning Analytics Platform. Proc. LAK2018, pp. 265-269, (2018).
4. N. Sclater, A. Peasgood, J. Mullan: Learning Analytics in Higher Education. A Review of UK and International Practice. Full Report. JISC, (2016).
5. N. Harada, M. Kimura, T. Yamamoto, Y. Miyake: System for Measuring Teacher-Student Communication in the Classroom Using Smartphone Accelerometer Sensors. HCI 2017 Interact. Context, pp. 309–318, (2017).
6. L. M. Schleifer, T. W. Spalding, S. E. Keric, J. R. Cram, R. Ley, B. D. Hatfield: Mental stress and trapezius muscle activation under psychomotor challenge: A focus on EMG gaps during computer work. Psychophys., v45, (2008).
7. N. L. Elwess, D. F. Vogt: Heart Rate and Stress in a College Setting. J. College Biol. Teaching, v31, n4, pp. 20–23, (2005).
8. D. A. Bligh, What's the Use of Lectures? San Francisco, CA: Jossey-Bass, (2000).
9. A. Kendon: Some functions of gaze-direction in social interaction. Acta Psychologica, v26, pp. 22-63, (1967).
10. H. H. Clark, T. B. Carlson: Hearers and Speech Acts. Language, v58, n2, pp. 332-373, (1982).
11. J. A. Hall, E. J Coats, L. S. LeBeau: Nonverbal Behavior and the Vertical Dimension of Social Relations: A Meta-Analysis. Psychol. Bull. v131, n6, pp. 898–924, (2005).
12. M. Zancanaro, B. Lepri, F. Pianesi: Automatic detection of group functional roles in face to face interactions. ICMI '06: Procs. 8th int. conf. Multimodal interfaces, pp.28–34, (2006).
13. F. Nihei, Y. I. Nakano, Y. Hayashi , H. Huang , S. Okada: Predicting Influential Statements in Group Discussions using Speech and Head Motion Information, ICMI '14: Procs. 16th Int. Conf. Multimodal Interaction, pp.136–143, (2014).
14. Kuniaki Yajima, Yoshihiro Takeichi, Jun Sato: Detecting Concentration Condition by Analysis System of Bio-signals for Effective Learning, Information and Communication Technology, pp 81–89, (2017).
15. S. Nomura, M. Hasegawa-Ohira, Y. Kurosawa, Y. Hanasaka, K. Yajima, and Y. Fukumura: SKIN TEMPRETURE AS A POSSIBLE INDICATOR OF STUDENT'S INVOLVEMENT IN ELEARNING SESSIONS, 2011NETs Int. Conf. Int. Studies, (2011).
16. S. Charoenpit, M. Ohkura: New E-learning System focusing on Emotional Aspects using Eye Tracking, Procs 5th Int. Conf. on Applied Human Factors and Ergonomics AHFE 2014, pp.6161–6170, (2014).
17. S. Sakuragi, Y. Sugiyama, K. Takeuchi: Effects of Laughing and Weeping on Mood and Heart Rate Variability, Journal of PHYSIOLOGICAL ANTHROPOLOGY and Applied Human Science, v21, n3, pp. 159–165, (2002).
18. I. Summers, T. Cofelt, R. E. Horton: Work-group cohesion. Psych. Rep., v63, pp. 627–636, (1988)
19. P. Ekman, Emotion in the Human Face. New York: Pergamon Press, (1972).
20. E. Garcia-Ceja, V. Osmani, O. Mayora: Automatic Stress Detection in Working Environments From Smartphones' Accelerometer Data: A First Step. IEEE J. Biomed. Health. Info., v20, n4,(2016).
21. M. Janidarmian, F. A. Roshan, K. Radecka, Z. Zilic: A Comprehensive Analysis on Wearable Accelerometers in Human Activity Recognition. Sensors (Basel), v17, n3, (2017).
22. Y. Okada, T. Y. Yoto, T. Suzuki, S. Sakuragawa, T. Sugiura: Wearable ECG Recorder with Accelerometers for Monitoring Daily Stress. Conf. Proc. IEEE Eng. Med. Biol. Soc., pp. 4718-4721, (2013).

23. J. Sztajzel: Heart-rate variability: a noninvasive electrocardiographic method to measure the autonomic nervous system. Swiss Med. Weekly, n134, pp. 514–522, (2004).
24. J. Wijsman, B. Grundlehner, H. Hermens: Trapezius Muscle EMG as Predictor of Mental Stress. ACM Trans. Embedded Comput. Sys, v12, n4, pp. 155–163, (2010).
25. S. Mallat, W. L. Hwang: Singularity detection and processing with wavelets. IEEE Trans. Inform. Theory, n38, pp. 617-643, (1992).
26. G. Zheng, Y. Wang, Y. Chen: Study of Stress Rules Based on HRV Features. J. Comput., v29, n5, pp. 41–51, (2018).
27. M. O. Mendez, J. Corthout, S. Van Huffel, M. Matteucci, T. Penzel, S. Certti, A. M. Bianchi: Automatic screening of obstructive sleep apnea from the ECG based on empirical mode decomposition and wavelet analysis, Physiol. Means., v31, pp. 273–289, (2010).