Quality and User Experience (2023) 8:10
https://doi.org/10.1007/s41233-023-00063-6

RESEARCH ARTICLE

Detection of nodding of interlocutors using a chair-shaped device and investigating relationship between a divergent thinking task and amount of nodding

Kento Nishimura[1] · Kodai Ito[2] · Ken Fujiwara[3] · Kazuyuki Fujita[4] · Yuichi Itoh[2]

Received: 31 May 2023 / Published online: 24 October 2023

Abstract
We evaluate a group's intellectual productivity in terms of its nodding. We first propose a method that detects nodding using a chair-shaped sensing device: SenseChair. Normalized time series of 3D data (i.e., the center-of-gravity [X, Y] and weight changes [W] on the seat) were submitted to a short-time frequency analysis with a Hanning window function. Nodding was detected by a neural network using the obtained short-time frequency data as features. We confirmed that this method's accuracy was comparable to that of an existing one that uses cameras. Next 13 groups of six speakers were engaged in a divergent thinking task where their nodding was detected by our proposed method. The results showed that the amount of nodding increased after idea generation, suggesting a positive relationship between the amount of nodding and the group's intellectual productivity. However, we found no significant correlation between the quality of each subjectively rated idea and the amount of nodding (i.e., the idea-level correlation). Therefore, we can conclude that our method was successful in detecting nodding from the seated participants as a behavior with functions of local coordination and agreement.

Keywords Sensing device · Intellectual productivity · Nonverbal behavior · Nodding

Introduction

Intellectual production activities that involve the cooperation of multiple people might yield results that exceed the capabilities of each individual alone. For instance, methods leveraging group creativity, such as brainstorming and hackathons, have been proposed. Since society demands innovation, much research has focused on revealing the antecedents of intellectual productivity in groups. For example, Woolley et al. report a psychometric methodology for quantifying a factor termed "collective intelligence," which reflects how well groups perform on a similarly diverse set of group problem-solving tasks. They quantitatively evaluated the intelligence of a group as collective intelligence, noting that groups that performed well on one cognitive task tended to perform well on others. Collective intelligence consists of the average social sensitivity of a group's members, the proportion of females in the group, and the equality of the distribution of conversational turn-taking Woolley

Kodai Ito, Ken Fujiwara, Kazuyuki Fujita and Yuichi Itoh have contributed equally to this work.

✉ Kodai Ito
kodai.ito@it.aoyama.ac.jp

Kento Nishimura
kento.nishimura@x-lab.team

Ken Fujiwara
k.fujiwara1985@gmail.com

Kazuyuki Fujita
k-fujita@riec.tohoku.ac.jp

Yuichi Itoh
itoh@it.aoyama.ac.jp

1 Graduate School of Information Science and Technology, Osaka University, 2-1, Yamadaoka, Suita, Osaka 5650871, Japan

2 College of Science and Engineering, Aoyama Gakuin University, 5-10-1, Fuchinobe, Chuo-ku, Sagamihara, Miyagi 2525258, Japan

3 Department of Psychology, National Chung Cheng University, No.168, Sec. 1, University Rd., Minhsiung 621301, Chiayi, Taiwan

4 Research Institute of Electrical Communication, Tohoku University, 6-6 Aoba, Aramaki, Aoba-ku, Sendai, Kanagawa 9808579, Japan

et al,[35]. Other nonverbal behaviors correlated to performance in intellectual production activities remain unclear. In the research field on social sensing, the use of sensing technologies to gather data related to human activities and behaviors in social contexts, assessing group interactions is an area of growing interest due to the prevalence of various types of sensing devices. Our study, which resembles such an approach, attempts to predict task performance in a group by sensing nonverbal behavior.

In this study, we propose SenseChair (Fig. 1), which is a chair-shaped device that detects nodding. SenseChair acquires the 3D time series (i.e., the center-of-gravity [X, Y] and the weight changes [W]) of a sitting person Tsuzuki et al, [31]. It is a simple and unobtrusive sensing approach because it only requires that its users are seated. The comfort afforded by SenseChair resembles a conventional office chair from which the user's psychological state can be sensed in an environment that closely resembles daily life. A user can stop being measured by simply standing up whenever she/he feels uncomfortable. SenseChair feature that addresses privacy concerns by avoiding the use of cameras or microphones, helps to mitigate privacy concerns related to visual or audio data capture. In our study, SenseChair simultaneously detects each interlocutor's nodding during a group activity.

For this paper, we conducted two experiments using SenseChairs. In the first, three participants engaged in the Consensus Game Hall, [9] and the Alternative Uses Task Guilford, [7] to see how accurately our device detected nodding. In the second experiment, we conducted the Alternative Uses Task with 13 groups of six participants and investigated the relationship between the amount of nodding detected using SenseChair and the ideas generated during the task. Based on these results, we compared the amount of nodding around the points of time when ideas were generated with the amount of nodding around the other points. This analysis approach allowed us to investigate the relationship between the amount of nodding and the group's intellectual productivity. We also focused on the subjective quality of ideas. Next we computed the idea-level correlation by investigating the association between the calculated questionnaire results and the amount of nodding. We also examined the group-level correlation. The subjective quality of each idea was averaged within a group to examine the relationship between the total amount of nodding and the idea quality generated through the task.

Related work

Social sensing with cameras

Numerous studies have been conducted to assess conversational situations by detecting nonverbal cues such as volume, gestures, and eye direction in conversational environments using cameras. Stiefelhagen et al. set a camera in a conversation space and modeled conversation structure by estimating the view directions Stiefelhagen et al, [28], and Otsuka et al. estimated speakers and listeners in conversations from view directions Otsuka et al, [24]. However, camera-based methods also have limitations such as occlusion, blind spots, and labor-intensive installation. Moreover, again perhaps the existence of sensors themselves might disrupt communication by adding psychological stress on speakers if user behaviors are obviously being sensed by them Won et al, [34]. Therefore, it is desirable to use a fine sensing method in an environment where the user is unaware of the sensor's presence as much as possible.

Social sensing without cameras

Methods that measure nonverbal information without cameras are mainly divided into two types: using wearable sensors or sensors placed in the environment. Wearable sensors make it possible to monitor users and situations through conversations regardless of activities. Sumi et al. proposed an IMADE room, where a microphone, an eye mark recorder, a motion sensor, and a heartbeat sensor are mounted on users who can move and communicate with each other. With various kinds of data, they investigated the structure of group

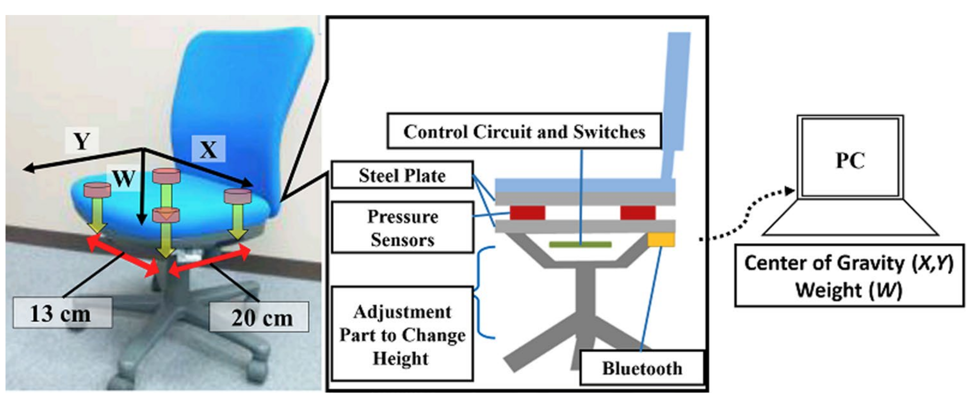

Fig. 1 Configuration of SenseChair and system overview

communication Sumi et al, [29]. Olguin et al. developed Sociometric Badges for which they implemented a microphone, an accelerator, and infrared sensors and evaluated users' daily communication ability Olguin-Olguin and Pentland, [22]. Zhang et al. used a wearable device with an accelerometer to calculate the features of daily activity and compared the characteristics within a group to assess its cohesion Zhang et al, [38]. Although these methods enable researchers to acquire rich information and analyze it in more detailed fashion, unfortunately, users must attach wearable sensors and/or devices. In addition, the presence of sensors often interferes with natural behavior of users.

On the other hand, methods that employ sensors that are located in their environments do not need extra time to mount them on user bodies before conversations. In research using a microphone, Kennedy et al. proposed a system that extracts the critical parts in conversations from voice tones Kennedy and Ellis, [13]. Wrede et al. proposed a method for detecting conversation excitement from speech Wrede and Shriberg [36]. A microphone enables us to estimate situations in conversations in more details based on the volume and pitch of user voices. However, estimating conversational situations in the absence of speech can be challenging, impairing its usefulness in situations where long periods of silence may occur. We cannot obtain useful information from a microphone in non-active conversation situations where the users are only listening to speech, such as conversations with stranger(s).

Chair-based sensing

Chairs are familiar elements in daily life and their presence has little stress to users. In settings like meetings or discussions, sitting in a chair to converse is extremely natural. Therefore, it is considered feasible to sense the everyday communication behaviors and conditions of users through chairs. Furthermore, this approach is non-invasive and non-wearable, making it highly convenient for everyday use. Consequently, chair-based sensing methods are thought to be highly effective for recognizing conversational situations.

Commonly, methods for recognizing the conditions of seated users involve embedding sensors in the chairs. Itoh has installed pressure sensors in car seats to estimate the position and state of the driver's feet [11]. In studies targeting chairs found in homes or offices, Tan and his team have proposed a method for identifying a user's seated posture based on the distribution of pressure on the seat and backrest [30]. Tan's group mainly targets the identification of user postures and has installed 2016 pressure sensors (42x48) on both the seating surface and backrest, achieving identification accuracy of 96% when individual posture data exists and 79% when using other people's posture data. Multu and others have identified postures similar to those used by Tan with fewer sensors-only 19 in total [21]. Vanhala and his team have placed ultra-sensitive pressure sensors known as ElectroMechanical Film (EMFi) in chairs to investigate how users' bodily movements change in response to the emotional expressions of computer agents [32].

Thus, it is considered that the distribution of pressure on the seat and backrest is effective for estimating the behavior and posture of users. However, there are challenges such as the trade-off between identification accuracy and the cost of using multiple pressure sensors. The pressure sensor needs to be spread over the entire seat surface, and its cost increases in proportion to the size of the seat surface.

Nodding

Such nonverbal behaviors as laughing, nodding, and gesturing by a conversation's listeners are called backchannels and play an important role in interpersonal communication Yngve, [37]. One form of backchannel is nodding. Nodding provides a variety of functions, including "indicating agreement" and "placing emphasis"Maynard, [19]. Nodding, a natural behavior that occurs frequently and unconsciously during conversations, can create a variety of positive impressions. For example, when a listener nods, a speaker tends to talk more Matarazzo et al, [17]; listeners also tend to agree more when they are nodding Briñol and Petty, [3]. Nodding creates a positive impression and engenders trust in a discussion Oshima, [23]. Furthermore, psychological research suggests that the frequency of nodding in face-to-face conversations may reveal individual characteristics or even predict communication outcomes; applicants who nod more frequently in job interviews are more likely to be hired than those who nod less frequently Gifford et al, [6], McGovern et al, [20]. The role of nodding has been especially scrutinized in communication among Japanese Kita and Ide, [15], the target nationality of our study. For example, Kogure demonstrated that silence during conversations is filled with mutual or simultaneous nodding to maintain a cooperative atmosphere during interactions Kogure, [16]. Although previous studies usually focused on dyadic interactions Hale et al, [8], nodding undoubtedly offers benefits even in group communication. Thus, we believe that it is a promising behavioral indicator that facilitates the efficacy of group interaction and improves group intellectual productivity. If virtual reality technology is employed, using a head-mounted display (HMD) to detect nodding is effective. Since the HMD is worn directly on the head, it ensures high reliability in nodding detection. Aburumman et al. investigated the relationship between nodding behavior and likability as well as reliability within a VR environment [1].

To detect nodding in conversational settings, an image-processing technique commonly captures the images of interlocutors. One method creates a 3D model of a speaker's face from video taken by a camera and detects nodding based on the model's rotation angle JINDAI et al, [12]. In this method, however, a picture must be taken from in front of the target, creating camera preparation/installation costs as well as adding stress on each participant. As pointed out by Won et al. [34], the mere fact that a behavior is being sensed undoubtedly causes anxiety in some speakers, ironically inhibiting smooth communication. Discreet sensing is required that resembles everyday life as closely as possible.

SenseChair

System design

We overcome these problems by proposing methods that acquire the body-sway data of users conversations using a chair-shaped device called SenseChair. Figure 1 shows its basic principle. SenseChair has four pressure sensors at each edge under its seat. These sensors have strain gauge type force sensor's. Since 50 kg is the pressure sensor's capacity, the pressure sensors must be maintained horizontally and at an equal height so that they don't exceed their capacity. We put four pressure sensors on each edge of the chair between two 5-mm thick, steel plates to keep the sensors horizontal and to equally disperse the user's weight (Fig. 1). The system can acquire both center-of-gravity and weight data in real time by a maximum frequency of 100 Hz. SenseChair's implementation is simpler than other chair-shaped devices since it has only four pressure sensors. These four sensors and the circuitry to control them can currently be purchased for less than 10 US dollars. Therefore, its installation and implementation costs are reasonable, and it is feasible for daily use since its implementation can be easily applied to other chairs with different shapes.

Center-of-gravity calculation

The center-of-gravity data include an X-coordinate, which is a left-right direction toward the front of the seat, and a Y-coordinate, which is a front-rear direction toward its front. It's important to note that what we are measuring here is not the force vector, but rather the information about where the seated user is placing their center of gravity on the seat. The weight data denote the vertical direction movements of users. Each sensor value is represented as FL (front left) [kg], FR (front right) [kg], BL (back left) [kg], BR (back right) [kg], and the distance between sensors to the X-axis direction is Lx [m] and Ly to the Y-axis direction. The center-of-gravity (X, Y) and the weight (W) are calculated as follows:

$$X = \frac{FR \times Lx + BR \times Lx}{FL + FR + BL + BR} \tag{1}$$

$$Y = \frac{FR \times Ly + FL \times Ly}{FL + FR + BL + BR} \tag{2}$$

$$W = FL + FR + BL + BR, \tag{3}$$

where BR is the origin of the coordinate axes, the X-axis positive direction is the left direction, and the Y-axis positive direction is the front direction toward the chair. The data from the SenseChair define the center-of-gravity as X [m] and Y [m] and the weight as W [kg].

Experiment 1

Our experiments were carried out according to the principles expressed in the Declaration of Helsinki and approved by our local University Research Ethics Review Committee. All participants provided written informed consent to join, and this consent was also approved by the ethics committee.

Experimental methods

We asked three participants to engage in the Consensus Game Hall, [9] and the Alternative Uses Task Guilford, [7] to see how accurately our device detected nodding.

Participants

Three male university students in their 20 s participated in the experiment. They were all acquaintances and in the same age. Since females are more likely to synchronize their behavior Fujiwara et al,[4], we chose the male group for reliability.

Task and conditions

The participants sat on three initialized SenseChairs, and after a brief explanation of the experiment, performed two teamwork tasks for 20 min each. The first was a Consensus Game that prioritized 15 tools for surviving in the desert. The participants individually thought about the situation for ten minutes, discussed it among themselves, and agreed on one answer for the whole team. The second game was the Alternative Uses Task for which they listed as many uses for a brick as possible. The participants verbally tabulated the uses without taking any notes or using information terminals during the discussion. We recorded the video with a webcam to label the data. SenseChair's sampling frequency was set to 30 Hz to match the sampling rate of the cameras.

Data acquisition

After completing the experiment, three annotators (two participants and an experiment supervisor) explicated the recorded video by their nods. In this annotation, we created correct answer data for nodding when at least two out of the three people judged an answer as nodding. One participant did not attend as an annotator; however, the remaining two participants, who engaged in face-to-face conversation with him, annotated the data. As a result, the annotated data are considered to be sufficiently reliable. The number of annotated nods was 110 for Subject 1, 120 for Subject 2, and 296 for Subject 3, totaling 526 nods. We used this dataset of 526 nods to perform nod detection and evaluated the degree of agreement between the nods identified by the system and the ground truth bythe F-measures. We used a neural network (NN) to detect nodding. Its structure and a method that used camera images as feature values are based on a Multi-Scale Convolution-LSTM Sharma et al, [27] that detected head movements with very high accuracy.

Next we describe a method that uses center-of-gravity and weight changes acquired by the SenseChairs as feature values. First, we standardized the time-series data of the center-of-gravity and weight, analyzed the frequency, and made them into a power spectra. We used a 64-point Hamming window as the window function and set the amount of time shift to one sample each. We used the power spectra of the frequencies in a range from 1.5 to 6.0 Hz. Based on previous studies, this range is considered well characterized because 96% of the time taken for a single nod is distributed from 0.17 to 0.57 s Kihara et al, [14]. Since the data acquired at 30 Hz were frequency analyzed in a 64-point Hamming window, there were 11 corresponding power spectra in each direction. 33 power spectra were acquired in the X, Y, and W directions, and we used six of the averages and variances of the data acquired in the windows: a total of 39 features. For the camera images, we extracted 28 facial landmarks as features using OpenFace Baltrušaitis et al, [2].

Results

We calculated the F-measures of the three participants using the camera detection and SenseChair methods where each method judged whether nodding occurred in each frame of the data acquired at a sampling frequency of 30 Hz. The results are listed in the left half of Table 1. The F-measures of participant A are high for both methods. However, for participant B, the F-measures of the camera method are much lower, probably because he wore glasses. In fact, previous studies Wall et al, [33] reported that the accuracy of head movement detection fell when participants wore glasses. The F-measures of participant B with the SenseChair method are also lower than those of participant A, but not as low as in the camera method, indicating a possible advantage of the SenseChair method. For participant C, the F-measures are not very high in either method. We believe that neither method accurately detected the nodding because participant C fidgeted during the experiment.

Next we calculated the accuracies of the methods to detect a series of nods as a single cluster since nodding is a continuous motion and it may not be reasonable to detect it on a frame-by-frame basis. First, we counted as one nodding cluster those nods whose intervals (between nods) were less than 0.30 s in the detected data. If the nodding clusters in the detected data overlapped the nodding clusters in the training data, even by one frame, or if they existed within 0.30 s, a nodding cluster was detected. We chose this value of 0.30 s because the time required for a single nod is distributed around 0.30 s. We would like to highlight that clustering can be seen as a means to remove noise by ignoring short-term fluctuations. We believe this technique is applicable even during the actual operation of the system.

The results of the cluster-by-cluster evaluation using this method are shown in the right half of Table 1. The F-measures of all the participants and the methods improved more when evaluated on a cluster-by-cluster basis than on a frame-by-frame basis. The F-measures of the participants and the methods with low F-measures greatly improved when evaluated on a cluster-by-cluster basis, suggesting that the frame-based detection's time resolution was too fine to a minor shift in the detection of nodding. Since such minor deviations were barely perceived by the speakers and only slightly impact daily conversations, we believe that cluster-based detection satisfies our objective. This result is the first evidence that the SenseChair method detects nodding with accuracy comparable to a camera method.

Table 1 Main results of Experiment 1 showing method and F-measure of each subject: (frame) indicates evaluation results in frames and (cluster) indicates evaluation results in clusters

Participant /Method	Camera (frame)	SenseChair (frame)	Camera (cluster)	SenseChair(cluster)
Participant A	0.795	0.771	0.800	0.834
Participant B	0.029	0.406	0.388	0.536
Participant C	0.224	0.086	0.354	0.426

Experiment 2

In this experiment, we conducted the Alternative Uses Task with 13 groups of six participants and investigated the relationship between the amount of nodding detected using SenseChair and the ideas generated during the task.

Experimental methods

Participants

Fifteen groups of six males in their 20 s to 40 s (90 people total) were recruited through a temporary employment agency to participate in the experiment. The age distribution was 75 in their 20 s, 8 in their 30 s, and 7 in their 40 s. Due to measurement errors caused by inadequate equipment and other factors, we only used the data from 13 of the 15 groups (78 participants).

Task and conditions

After informing the participants of our experimental intentions, we recorded their actions with a 360-degree camera in addition to the time-series center-of-gravity and weight data by the SenseChairs. The participants sat on initialized SenseChairs, and after a brief explanation, conducted the Alternative Uses Task. They verbally listed as many uses for bricks as possible without taking notes or recording anything on information terminals. Since communicating persons never generate higher frequency of movements than those from trembling, which is a human physiological phenomenon that generally occurs at 10 Hz Sakamoto et al, [26], the sampling frequency of the SenseChairs was set to 20 Hz.

Data acquisition

We used SenseChairs to detect nodding as in Experiment 1. We calculated the amount of nodding using the frame-by-frame nodding data. Note that this frame-by-frame nodding data was corrected by the clustering process described in Sect. 4.3. First, we assigned a nod value to each frame for each participant. The amount of nodding a_{kt} for participant k at certain frame t was calculated by the following equation, where S_k is the total number of frames in which participant k nods:

$$a_{kt} = \begin{cases} \frac{1}{S_k}\,(k \textit{ is nodding at frame } t.) \\ 0\,(k \textit{ is not nodding at frame } t.) \end{cases} \tag{4}$$

By summing up the amount of nodding assigned to each participant, we calculated the number of nods for that frame. Nodding amount A_t for a group of six people at frame t is calculated by the following formula:

$$A_t = \sum_{k=1}^{6} a_{kt}. \tag{5}$$

In other words, we weighted each participant by the reciprocal of the total number of nodding frames. Those who nodded frequently had a small effect on the calculation of the nodding amount; those who nodded infrequently had a larger effect. We conducted the following analysis based on nodding amounts calculated in this way.

Results

Analysis 1

In this analysis, we investigated the differences in the amount of nodding between around a point of time when the idea was generated (where using a brick was mentioned) and the other points. We defined "around" as five seconds before and after each point of time. We defined the point of time when the idea was generated as the timing when the participants started discussing the idea. In addition, we focused on the point after the idea generation and also investigated the point before the idea was generated. The set time should be longer than the idea's duration. However, if it is too long, perhaps nodding will be included that was induced by something unrelated to the idea. When we examined the duration of all the ideas generated during the task, 95% were less than 4.91 s. Therefore, we adopted a duration of five seconds, which exceeds the duration of most of the ideas.

Next we selected the points of time generated by the idea and the points to be compared. To avoid any bias, we extracted the points to be compared from the task's entire time. We first defined the frames that discussed ideas as "idea frames" and frames that did not discuss them as "non-idea frames." To observe the differences between with/without conversations, we classified them into two groups: "conversation frame" and "non-conversation frame." The frames do not overlap. If several ideas occurred in 5 s, only the first one was counted.

We divided the time-series data of the group activity into three segments: A) where an idea was generated (conversation segment with an idea), B) where no idea was generated but there was a conversation (conversation segment without an idea), and C) where there was no conversation (non-conversation segment). For comparison, these three segments were further divided into the amount of nodding five seconds before (before) and after (after) for a total of six segments. Table 2 shows the classification of the segments. Figure 2 compares the six segments,

Table 2 Classification of segments

Segment	Definition
A	Conversation segment without idea (before)
B	Non-conversation segment (before)
C	Conversation segment with idea (before)
D	Conversation segment without idea (after)
E	Non-conversation segment (after)
F	Conversation segment with idea (after)

where a violin plot was overlaid with a box-and-whisker diagram of the amount of nodding in each segment. Since this study focuses on comparing the amount of nodding between the points of time when an idea is generated and other points, we compared the amount of nodding calculated for the same section (before or after) for the "conversation segment with an idea" and two other segments using the Brunner-Munzel Test, which is a non-parametric test of the significant differences between two groups. We focused on the "conversation segment with an idea" and compared the amount of nodding five seconds before and after the segment. Given a total of five multiple comparisons, the level of significance was adjusted using the Bonferroni correction: $.05/5 = .01$. The results show that when comparing the amount of nodding for five seconds before each segment (i.e., comparison among A, B, and C in Fig. 2), it was significantly different between the segments: $A > C(p < .001)$, $B < C(p < .001)$. In contrast, when comparing the amount of nodding five seconds after each segment (i.e., comparison among D, E, and F in Fig. 2), it was significantly higher in the "F: conversation segment with idea (after);" $D < F(p < .001)$, $E < F(p < .001)$. When comparing "C: conversation segment with idea (before)" with "F: conversation segment with idea (after)," the amount of nodding in the latter segment was significantly higher ($p < .001$). These results suggest that nodding is more likely to occur after idea generation.

Analysis 2

In this analysis, we investigated the relationship between the amount of nodding after idea generation and the idea's quality. Since the results of Analysis 1 showed that nodding after idea generation was a dominant pattern, we calculated the amount of nodding after idea generation by summing the amount of nodding in the first five seconds after idea generation, as in Experiment 1. We quantified the quality of the ideas through a questionnaire survey. 428 ideas were generated during the task, and after removing duplicate and synonymous ideas, 211 different ideas were analyzed. We used the following four items, which captured the multifaceted nature of the outcomes in the divergent thinking task Hennessey and Amabile, [10], Plucker et al, [25] to evaluate them: (Q1) *Is this idea impressive?* (Q2) *Is this idea original?* (Q3) *Is this idea convincing?* and (Q4) *Is this idea socially acceptable?* We evaluated the questionnaire on a 5-point scale: *(5: greatly agree; 4: agree; 3: neither; 2: disagree; and 1: strongly disagree)*. These four items were chosen not to measure the overall creativity with a single item but to test whether items related to creativity (even in the opposite direction) collectively show results in the same direction. We administered questionnaires to 11 people and

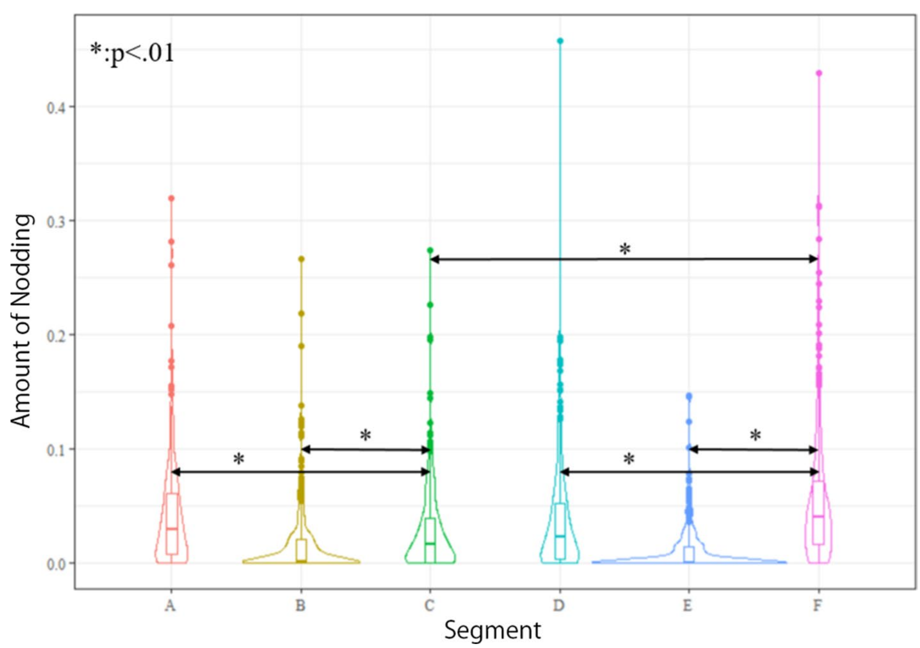

Fig. 2 Main results of Analysis 1 showing amount of nodding for six segments in a violin plot overlaid with a box-and-whisker diagram: Segments A and C, B and C, D and E, E and F, and C and F were compared using Brunner-Munzel Test. Level of significance was adjusted using Bonferroni correction ($p = .01$)

calculated the average results for each item and each idea. We conducted a correlation analysis using Spearman's rank correlation coefficient between the questionnaire results and the amount of nodding and conducted a multiple regression analysis using the questionnaire results as explanatory variables and the amount of nodding as an objective variable.

Analysis 2a: creativity at idea level Fig. 3 shows a scatter plot of the ideas generated during the task. The horizontal axis shows the amount of nodding five seconds after idea generation and the vertical axis shows the questionnaire results on the subjective quality of the ideas for each (a) to (d). As in Fig. 3, the correlation (Spearman's rank correlation) was small and non-significant except for Q2; $Q1 : -.083(p = .084)$, $Q2 : -.100(p = .036)$, $Q3 : .056(p = .242)$, and $Q4 : .091(p = .056)$. Thus, we found no idea-level correlation between the amount of nodding and the subjective quality of each idea.

Analysis 2b: creativity at group level Fig. 4 shows a scatter plot of the total frames detected as nodding in the group and the averaged quality of the ideas during the task. The horizontal axis shows the amount of nodding five seconds after idea generation, and the vertical axis shows the questionnaire results on the subjective quality of the ideas for each (a) to (d). For each variable, the correlation coefficient (Spearman's rank correlation) was not small; $Q1 : .489(p = .090)$, $Q2 : .324(p = .280)$, $Q3 : -.319(p = .289)$, and $Q4 : -.357(p = .231)$. Therefore, it can be said that we were able to detect nodding as an action that serves the simple functions of local coordination and agreement, rather than focusing on the quality of the ideas or group creativity.

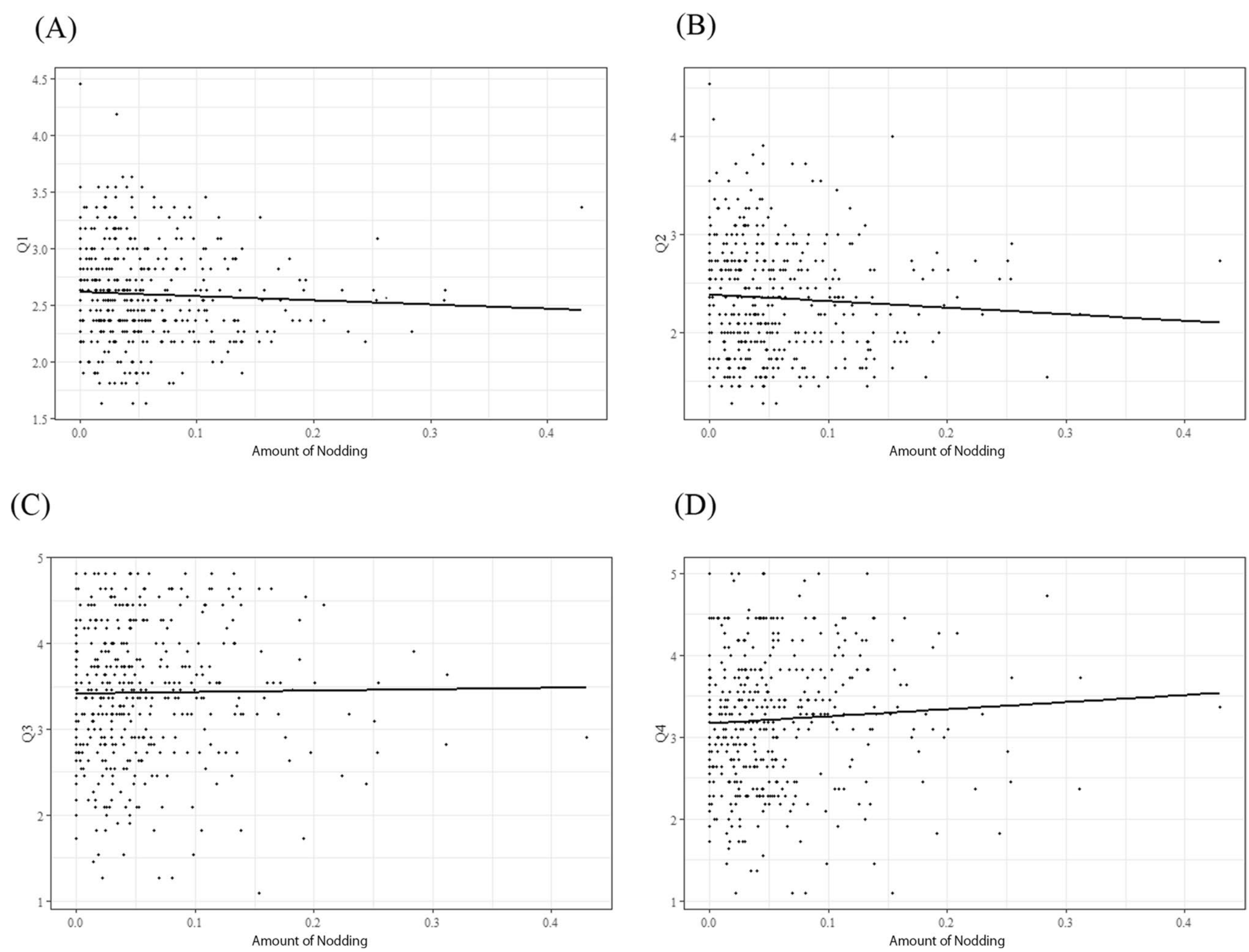

Fig. 3 Main results of Analysis 2a showing amount of nodding within five seconds after idea generation (horizontal axis) and questionnaire result for each (**A**) to (**D**) (vertical axis): Questionnaire items are (**A**) Q1: impressive, (**B**) Q2: original, (**C**) Q3: convincing, and (**D**) Q4: socially acceptable

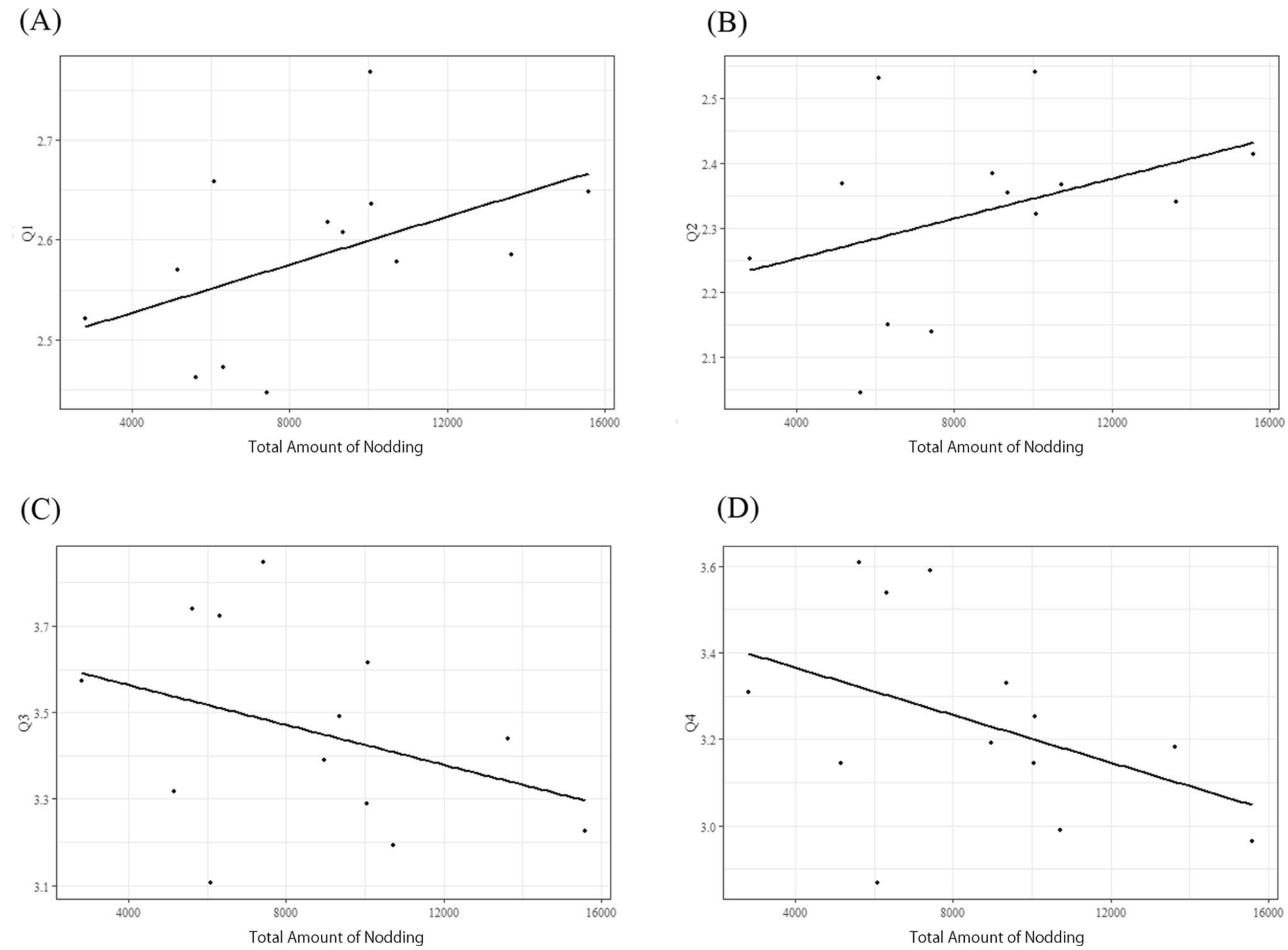

Fig. 4 Main results of Analysis 2b showing total amount of nodding within five seconds after idea generation (horizontal axis) and questionnaire result for each (**A**) to (**D**) (vertical axis): Questionnaire items are (**A**) Q1: impressive, (**B**) Q2: original, (**C**) Q3: convincing, and (**D**) Q4: socially acceptable

Discussion

We proposed a nodding-detection method using the SenseChair in a group task. In Experiment 1, we compared the detection accuracy of the proposed and existing methods using cameras and confirmed that both methods were comparable in term of detection accuracy. In Experiment 2, the amount of nodding increased more after an idea was generated compared to before an idea was generated or other points (Analysis 1). However, we found no significant correlation between the amount of nodding after idea generation and the subjectively rated creativity at either the idea level (Analysis 2a) or the group level (Analysis 2b).

Since our proposed nodding-detection method using the SenseChair was as accurate as existing methods using cameras, it seems to possess at least two advantages. One is cost. To detect nodding using a camera, a picture must be taken from in front of the target, which requires a camera for every interlocutor. It also puts a mental strain on the interlocutors who are being filmed Won et al, [34]. Whether online or offline, not everyone feels comfortable being filmed during group work. In this regard, because the SenseChair resembles an office chair and just measures the 3D information on the seat, users might feel more comfortable during its measurements. The second advantage concerns privacy, which we must protect. A facial photo captures more information than is necessary to detect nodding. The information processed in the SenseChair is simple but powerful enough to detect nodding. Even when filming is discouraged (e.g., a meeting involving confidential information), the SenseChair may be applicable.

Analysis 1 of Experiment 2 shows that the amount of nodding increased more after an idea was generated than before it was generated or other timings. These results were robust even after weighting the nodding frequency (i.e., nodding by those who nodded frequently as well as those who nodded infrequently). Thus, nodding functions as a reaction that flattens subsequent interactions. However, note that the ideas followed by much nodding did not necessarily denote high quality (Analysis 2a). Nodding may not be a direct indicator of a group's intellectual productivity, although it will indirectly reflect whether its communication is productive. Even if nodding served primarily as a simple local coordination and agreement, it is likely indicative of a group in which communication was successful.

Limitations

We developed SenseChair to measure center-of-gravity changes and proposed a method that automatically detected nodding. Although SenseChair offers several advantages, it requires more cost and space than a camera and may be affected by such unconscious actions as fidgeting. In our experiment, we placed cameras in front of participants to annotate or verify nodding. This decision may have biased their behavior, causing them to behave differently from their more typical behavior. In this study, we performed experiments using the Alternative Uses Task and analyzed the data under the premise that nodding is related to the generation of ideas. If the context were to change to free discussion, further experiments and analysis would be needed to examine whether the nodding gestures are influenced by simple agreement. Furthermore, the instructions for this study specified that participants should remain seated during the discussion, which may have led to behaviors that differ slightly from those in actual discussions. For instance, in such discussions, participants might stand up or use tools like a whiteboard. While chair-based sensors are inherently unable to sense users who are standing, incorporating analysis of movements like standing and sitting could offer new insights.

In the analysis, we focused on the quantity of nods and not their quality, such as amplitude and frequency. For example, previous studies argued that communication is smoother when body movements are synchronized in a certain frequency range (0.5−1.5 Hz) Fujiwara et al, [5]. The relationship between an idea and nodding may differ between sincere nodding and nodding while laughing at an idea's infeasibility. In fact, in this experiment, sometimes the participants nodded while laughing at relatively silly ideas, which might explain why no correlation was found in Analysis 2a. By quantifying the quality of the nods in terms of their amplitude and frequency and conducting further research, we believe that we can evaluate the quality of ideas based on speaker nods, i.e., judging a group's intellectual productivity. In this paper, since we did not deeply consider who nodded, most likely a leader's nodding will function differently than a follower's nodding. In other words, perhaps we should pay attention to the nods of individual participants and give them their own weight for evaluating a group's intellectual productivity. In addition, we calculated the idea-level and group-level correlations by investigating the relationships between the calculated questionnaire results and the amount of nodding. We used the proposed method in Analysis 1 to calculate the amount of nodding. Therefore, the proposed method's accuracy may have affected the results of Analysis 2. For example, we recognize that the results may be influenced by factors such as whether the subjects are wearing glasses and by biases in the attributes of the subjects, as the data used for training the nod detection were from only three individuals. To create a more generalized system, data should be collected from a larger number of subjects.

All the participants in this experiment were Japanese, and the majority were young males. It is possible that the timing and frequency of nods during conversations, or even the nodding motions themselves, could vary due to differences in age groups or cultural backgrounds between countries. While the results of this system may not be universally applicable, we have demonstrated the potential to estimate nodding behavior and conversational productivity by collecting data from subjects with diverse cultural backgrounds.

Conclusion

We conducted two experiments in which we asked three participants to engage in the Consensus Game Hall, [9] and the Alternative Uses Task Matsui and Hikono, [18] to determine how accurately our device detected nodding. The Consensus Game was employed because it is likely to yield a high number of nodding gestures, which serve as indicators of agreement. Meanwhile, the Alternative Uses Task was employed because it allows us to measure intellectual productivity in terms of idea generation. First, participants performed two tasks for ten minutes each, which were recorded using SenseChair and three webcams placed in front of each participant. Next we created a grand-truth label for the nodding timing to train the neural network. We prepared features such as changes in center-of-gravity and weight recorded using SenseChair and facial landmarks extracted from video data captured by cameras. We then compared their detection accuracy.

In the second experiment, we conducted the Alternative Uses Task with 13 groups of six participants and investigated the relationship between the amount of nodding detected using SenseChair and the ideas generated during the task. The group interaction data are based on Matsui et al. Matsui and Hikono, [18]. The participants sat in SenseChairs and engaged in the following 15-minute task. For the amount of nodding in a group, we adjusted the weights so that as more participants nodded throughout the task, the weight became smaller. This adjustment allows us to consider the nodding of those who do so infrequently. We calculated the number of group members who nodded to sum each weighted quantity of nodding and investigated the relationship between the amount of nodding and the ideas generated through the following three analyses. In Analysis 1, we compared the amount of nodding around the points of time when ideas were generated (where using a brick was mentioned) with the amount of nodding around the other points. This analysis approach allowed us to investigate the relationship between the amount of nodding and the group's intellectual productivity. In Analysis 2, we focused on the subjective quality of ideas. The questionnaire included the following four categories on a 5-point scale: impressive, original, convincing, and socially acceptable. Eleven people evaluated the ideas, which were averaged for each item and for each idea. As in Analysis 2a, we next computed the idea-level correlation by investigating the correlation between the calculated questionnaire results and the amount of nodding. In Analysis 2b, we examined the group-level correlation. The subjective quality of each idea was averaged within a group to examine the relationship between the total amount of nodding and idea quality through the task.

In the future, we will quantify nodding quality in terms of its amplitude and frequency because we believe that we can evaluate the quality of ideas based on speaker nods, i.e., evaluating a group's intellectual productivity.

Acknowledgements This work was supported in part by a grant from the Ministry of Education, Culture, Sports, Science and Technology of Japan (MEXT) to support the Research Center for the Realization of Society 5.0 and JSPS Grant-in-Aid for Scientific Research 16H02891. The authors express their profound gratitude to the Nuclear Safety Systems Research Institute, Inc. for providing the experimental data in this study.

Declarations

Conflict of interest The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Aburumman N, Gillies M, Ward JA et al (2022) Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions. Int J Hum Comput Stud 164:102819. https://doi.org/10.1016/j.ijhcs.2022.102819
2. Baltruvsaitis T, Robinson P, Morency LP (2016) Openface: an open source facial behavior analysis toolkit. In: 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), pp 1–10, https://doi.org/10.1109/WACV.2016.7477553
3. Briñol P, Petty RE (2003) Overt head movements and persuasion: a self-validation analysis. J Pers Soc Psychol 84(6):1123
4. Fujiwara K, Kimura M, Daibo I (2019) Gender differences in synchrony: females in sync during unstructured dyadic conversation. Eur J Soc Psychol 49(5):1042–1054. https://doi.org/10.1002/ejsp.2587
5. Fujiwara K, Kimura M, Daibo I (2020) Rhythmic features of movement synchrony for bonding individuals in dyadic interaction. J Nonverbal Behav 44(1):173–193. https://doi.org/10.1007/s10919-019-00315-0
6. Gifford R, Ng CF, Wilkinson M (1985) Nonverbal cues in the employment interview: links between applicant qualities and interviewer judgments. J Appl Psychol 70:729–736. https://doi.org/10.1037/0021-9010.70.4.729
7. Guilford J (1967) The nature of human intelligence. The nature of human intelligence, McGraw-Hill, New York, NY, US
8. Hale J, Ward JA, Buccheri F et al (2020) Are you on my wavelength? interpersonal coordination in dyadic conversations. J Nonverbal Behav 44(1):63–83. https://doi.org/10.1007/s10919-019-00320-3
9. Hall J (1971) Decisions, decisions, decisions. Psychol Today 5(1):51–54
10. Hennessey BA, Amabile TM (2010) Creativity. Annu Rev Psychol 61(1):569–598. https://doi.org/10.1146/annurev.psych.093008.10041
11. Itoh M (2008) Human monitoring-based driving support. IFAC Proceedings Volumes 41(2):15076–15087. , 17th IFAC World Congresshttps://doi.org/10.3182/20080706-5-KR-1001.02551
12. Jindai M, Nakamura K, Watanabe T (2016) A nodding detection system based on the active appearance model. J Adv Mech Des Syst Manuf. https://doi.org/10.1299/jamdsm.2016jamdsm0021
13. Kennedy L, Ellis D (2003) Pitch-based emphasis detection for characterization of meeting recordings. In: 2003 IEEE Workshop on Automatic Speech Recognition and Understanding (IEEE Cat. No.03EX721), pp 243–248, https://doi.org/10.1109/ASRU.2003.1318448
14. Kihara H, Fukushima S, Naemura T (2016) Analysis of human nodding behavior during group work for designing nodding robots. In: Proceedings of the 2016 ACM international conference on supporting group work. association for computing machinery, New York, NY, USA, GROUP '16, pp 433-436, https://doi.org/10.1145/2957276.2996287
15. Kita S, Ide S (2007) Nodding, aizuchi, and final particles in japanese conversation: How conversation reflects the ideology of communication and social relationships. J Pragmat 39(7):1242–1254. https://doi.org/10.1016/j.pragma.2007.02.009
16. Kogure M (2007) Nodding and smiling in silence during the loop sequence of backchannels in japanese conversation. Journal of Pragmatics 39(7):1275–1289. , nodding, Aizuchi, and Final Particles in Japanese Conversation https://doi.org/10.1016/j.pragma.2007.02.011
17. Matarazzo JD, Saslow G, Wiens AN, et al (1964) Interviewer head nodding and interviewee speech durations. Psychotherapy: Theory, Research & Practice 1:54–63. , place: US Publisher: Division of Psychotherapy (29), American Psychological Association https://doi.org/10.1037/h0088571
18. Matsui Y, Hikono M (2018) Relationship between performance and physical synchrony of the members in small group discussion: Focusing on high-synchrony groups. Ergonomics (in Japanese) 54(Supplement):2C2–4
19. Maynard S (1989) Japanese Conversation: Self-contextualization Through Structure and Interactional Management. Advances in discourse processes, Ablex Pub
20. McGovern TV, Jones BW, Morris SE (1979) Comparison of professional versus student ratings of job interviewee behavior. J Counseling Psychol 26:176–179. https://doi.org/10.1037/0022-0167.26.2.176
21. Mutlu B, Krause A, Forlizzi J, et al (2007) Robust, low-cost, non-intrusive sensing and recognition of seated postures. In: Proceedings of the 20th Annual ACM Symposium on User Interface

Software and Technology. Association for Computing Machinery, New York, NY, USA, UIST '07, p 149-158, https://doi.org/10.1145/1294211.1294237
22. Olguin-Olguin D, Pentland A (2010) Sensor-based organisational design and engineering. Int J Organ Des Eng 1(1–2):69–97. https://doi.org/10.1504/IJODE.2010.035187
23. Oshima S (2014) Achieving consensus through professionalized head nods: the role of nodding in service encounters in japan. Int J Bus Commun 51:31–57. https://doi.org/10.1177/2329488413516207
24. Otsuka K, Sawada H, Yamato J (2007) Automatic inference of cross-modal nonverbal interactions in multiparty conversations: "who responds to whom, when, and how?" from gaze, head gestures, and utterances. In: Proceedings of the 9th International Conference on Multimodal Interfaces. Association for Computing Machinery, New York, NY, USA, ICMI '07, p 255-262, https://doi.org/10.1145/1322192.1322237
25. Plucker JA, Karwowski M, Kaufman JC (2020) Intelligence and Creativity, 2nd edn., Cambridge University Press, Cambridge Handbooks in Psychology, pp 1087-1105
26. Sakamoto K, Shimizu Y, Mito K et al (2009) Tremor and vibratory perception in a living body - Functional evaluation of mechanical vibration -. Tokyo Denki University Press, Toyko, Japan
27. Sharma M, Ahmetovic D, Jeni LA, et al (2018) Recognizing visual signatures of spontaneous head gestures. In: 2018 IEEE Winter conference on applications of computer vision (WACV), pp 400–408, https://doi.org/10.1109/WACV.2018.00050
28. Stiefelhagen R, Yang J, Waibel A (2002) Modeling focus of attention for meeting indexing based on multiple cues. IEEE Trans Neural Netw 13(4):928–938. https://doi.org/10.1109/TNN.2002.1021893
29. Sumi Y, Yano M, Nishida T (2010) Analysis environment of conversational structure with nonverbal multimodal data. In: international conference on multimodal interfaces and the workshop on machine learning for multimodal interaction. Association for Computing Machinery, New York, NY, USA, ICMI-MLMI '10, https://doi.org/10.1145/1891903.1891958
30. Tan H, Slivovsky L, Pentland A (2001) A sensing chair using pressure distribution sensors. IEEE/ASME Trans Mechatron 6(3):261–268. https://doi.org/10.1109/3516.951364
31. Tsuzuki T, Itoh Y, Fujiwara K et al (2017) A method for sensing synchrony between communicating persons by sensechair and the evaluation toward conversation. Trans Human Interface Soc (in Japanese) 19(2):151–162
32. Vanhala T, Surakka V, Anttonen J (2008) Measuring bodily responses to virtual faces with a pressure sensitive chair. In: Proceedings of the 5th Nordic Conference on Human-Computer Interaction: Building Bridges. Association for Computing Machinery, New York, NY, USA, NordiCHI '08, p 555-559, https://doi.org/10.1145/1463160.1463243, https://doi.org/10.1145/1463160.1463243
33. Wall E, Schillingmann L, Kummert F (2017) Online nod detection in human-robot interaction. In: 2017 26th IEEE international symposium on robot and human interactive communication (RO-MAN), pp 811–817, https://doi.org/10.1109/ROMAN.2017.8172396
34. Won AS, Bailenson JN, Stathatos SC et al (2014) Automatically detected nonverbal behavior predicts creativity in collaborating dyads. J Nonverbal Behav 38(3):389–408. https://doi.org/10.1007/s10919-014-0186-0
35. Woolley AW, Chabris CF, Pentland A et al (2010) Evidence for a collective intelligence factor in the performance of human groups. Science 330(6004):686–688. https://doi.org/10.1126/science.1193147
36. Wrede B, Shriberg E (2003) Spotting "hot spots" in meetings: human judgments and prosodic cues. In: Proc. 8th European Conference on Speech Communication and Technology (Eurospeech 2003), pp 2805–2808, https://doi.org/10.21437/Eurospeech.2003-747
37. Yngve VH (1970) On getting a word in edgewise. Papers from the sixth regional meeting Chicago Linguistic Society, April 16-18, 1970, Chicago Linguistic Society, Chicago pp 567–578. https://cir.nii.ac.jp/crid/1571417124968122496
38. Zhang Y, Olenick J, Chang CH et al (2018) Teamsense: Assessing personal affect and group cohesion in small teams through dyadic interaction and behavior analysis with wearable sensors. Proc ACM Interact Mob Wearable Ubiquitous Technol. https://doi.org/10.1145/3264960