Conference State Estimation by Biosignal Processing — Observation of Heart Rate Resonance —

Masamichi Hosoda **Akira Nakayama** **Minoru Kobayashi** **Satoshi Iwaki**
NTT Cyber Space Laboratories, NTT Corporation
1-1 Hikarinooka, Yokosuka-shi, Kanagawa, 239-0847 Japan
hosoda.masamichi@lab.ntt.co.jp

Abstract
This paper discusses a conference state estimation method that uses only biosignal processing; linguistic understanding is avoided. In conventional dialogue communication research, the physiological characteristic of "entrainment" between the participants has been already reported. We extend "entrainment" to introduce the concept of "resonance" to grasp the relationship between attendees. We then propose a method which estimates the conference state from "resonance". First, using a multimedia conference system, we record conference data including time-series records of the participants' heart rates. By observing and analyzing the conference data, we assess the "resonance" phenomenon among the talkers. Using the "resonance correlation matrix", a newly proposed index based on the correlation of the heart rate data, conference participative state can be successfully estimated.

Categories & Subject Descriptors: H.1.2 [**Information Systems**]: User/Machine Systems – *Human information processing*

General Terms: Human Factors; Measurement; Theory

Keywords: Conference state; Biosignal; Heart rate; Resonance correlation matrix

INTRODUCTION

The traditional means of reviewing the proceedings of an actual conference has been to read the minutes. The text record allows us to discover what was said. For the minutes writer, however, it is difficult to fully participate to the discussion because he or she must concentrate on making the memo throughout the entire meeting. On the other hand, in the case of multimedia conferences, which have become popular due to the spread of broadband Internet connection, all conference contents including audio and visual data can be automatically recorded.

To investigate the basic problems and potential of the conventional Internet conference system, we conducted a telecommuting experiment by connecting multiple homes and several offices with a multi-point Internet conference system for about a year. In that experiment, users who couldn't attend a certain meeting were provided with archived video material of the meeting to review the proceedings. Contrary to our expectations, we found that such archives were not the best solution, mainly because it is too time consuming to carefully review a complete meeting from beginning to end.

CHI 2004, April 24–29, 2004, Vienna, Austria.
ACM 1-58113-703-6/04/0004.

Questionnaires on the experiment showed that there is strong demand for a more useful conference archive browsing tool. To quickly browse a conference archive, it is important to automatically estimate and annotate the conference status or structure such as topic boundaries, highlights, and the participative state. Such semantic information can be basically be derived from a linguistic understanding of what was said. The weak performance of current voice-recognition technologies, however, means that this requirement cannot be satisfied for conference conversations.

This paper, therefore, develops a method that can estimate participative state (the discussion is reaching agreement or not, etc.) from biosignal information. Conventional human communication research includes reports about "entrainment": the phenomenon in which one's biosignals like breath and heart rate become entrained and synchronized to those of another participant [4]. We extend this "entrainment" to introduce the concept of "resonance" to grasp the relationship between attendees. Using the "resonance correlation matrix", a newly proposed index based on the correlation of the heart rate, we try to estimate the conference participative state.

In the next section, we review existing research on communication analysis by biosignal processing, explain resonance, and propose a method to estimate conference state by utilizing resonance. Next, we describe the multimedia conference system used to record an actual conference, how the biosignal was captured/processed, and the calculation of the resonance correlation matrix. Finally, we explain the results of our experiment, and the relationship between the biosignal and conference status as discerned in the experiment.

COMMUNICATION ANALYSIS BY BIOSIGNAL PROCESSING

Conventional Research

Several papers have examined communication analysis by biosignal processing.

One approach centers on "entrainment": the synchronization of biorhythms like breath and heart rate while people are communicating [4]. Some reports state that communication can be enhanced by non-verbal interaction including the entrainment of body action rhythms and biosignal rhythms like breath and heart rate. During the course of the communication session these rhythms can change such that several users become synchronized. Other research tried to utilize this "entrainment" in a positive manner. They set many avatars in a virtual space, triggered avatar movement by voice, and generated body action entrainment to activate and support communication [5]. However, their study of body action entrainment among many members did not consider the use of biosignals. Only voice was used and no biosignals were collected or used by the system. In addition, they made no attempt to estimate the communication status of each scene from biosignal entrainment.

Another study utilized electroencephalograms as the biosignal to analyze conferences [1]. They calculated a value that represented the attendees' level of mental activity from electroencephalograms, determined their thinking status and used it as a conference index. Unfortunately, they concentrated on individual thoughts only, and didn't consider the relationship among attendees. This makes it difficult to apply this method to the estimation of conference status, which is determined by the relationship among attendees, for example agreement or disagreement.

Resonance and Conference Status

In this research, we extend the scope of "entrainment", and make the assumption that "resonance" captures not only the status of individual thoughts, but also the relationship among multiple attendees. We then propose a conference state estimation method that uses resonance.

Assumption of Resonance

The human body can be regarded as an independent system exhibiting autonomic biosignals, like heart pulses, that exhibit some "jitter" [3]. We assume that these systems are connected via the communication channel such that biosignal entrainment is possible. If the attendees can discuss the conference topic as equals, their psychological distance will change as the conference proceeds due to their differences in stance or opinion, and their association will strengthen or weaken. When the association strengthens, entrainment occurs and one or more of the biosignals synchronize. When the association weakens, neither entrainment nor biosignal synchronization occurs. The closeness/alienation of the association during a conference exhibits a multiple-to-multiple relationship through which the attendees exert an influence upon each other and the entrainment status changes widely. We extend biosignal "entrainment" and define biosignal "resonance" as the phenomenon of synchronization of a biosignal due to entrainment, which changes over time among the multiple relationships. This concept is shown in Figure1.

Proposal of Conference State Estimation Method

In this research, we propose to observe biosignal resonance to estimate the relationship among attendees and conference state.

RESONANCE OBSERVATION SYSTEM

Overview of System

Figure 2 shows the configuration of the multimedia conference system [2] used in this research. The multipoint AV communication server and the event record server were set up to establish and control a multimedia conference. They were connected to 4 client PCs via Gigabit Ethernet. Each client PC was connected to a USB camera, display, keyboard, mouse and headset. Each PC was set in a noise-free room so that no auditory/visual connection was possible between the 4 attendees and external noise was shut out. A typical client PC display screen is shown in Figure 3. An overview of system performance is shown in Table 1. Each attendee wore a biological sensor; the biosignals were logged by the biosignal collection server.

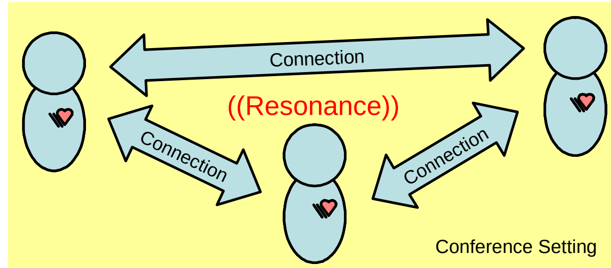

Figure 1. Concept of Resonance

Biosignal

Biosignal Measurement

We measured the heart pulse in this experiment since the equipment is relatively easy to use and its output is easy to log. We use the photoelectric pulse detection sensor widely used in areas such as sport training. A sensor was clipped to the ear lobe of each attendee. Figure 4 shows the details of the sensor. This sensor outputs the pulse wave peak interval (resolution of 10 ms), which is almost same as the interval between the R waves (RRI) captured in electrocardiograms.

Biosignal Processing

Heart Rate (HR) can be obtained by converting the measured RRI to the frequency of heart beat per minute. The variability of heart rate is called Heart Rate Variability (HRV). It is known that HRV is affected by 2 components: that derived from breathing (lies around 0.2 Hz – 0.3 Hz), and that due to blood pressure (lies around 0.1 Hz) [3]. To clearly measure resonance, we excluded these 2 components; we first obtained uniform interval data by sampling at the frequency of 100 Hz by linearly interpolating the instant heart rate, then extracted the component under 0.07Hz by a 9223rd order FIR linear phase low-pass filter. Figure 5 shows the example of HR data processing. "Original" is the frequency of heart beat per minute extracted from the sensor-output data. "Linear interpolation" is the result of linear interpolation. "FIR LPF" is the result of low-pass filtering.

Item	Performance
Face image frame rate	15 fps
Face image frame size	160 * 120
Audio sampling rate	22050 Hz
Audio delay approximately	220 ms
Simultaneous attendees	4
Recordable Event	Utterance start/end time Chat sends time etc.
Event record precision	ms unit

Table 1. Overview of System Performance

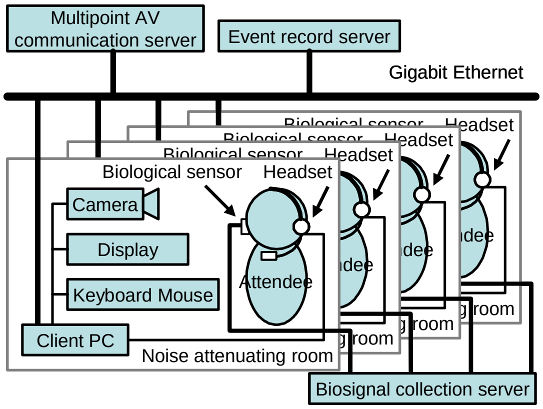

Figure 2. System Configuration

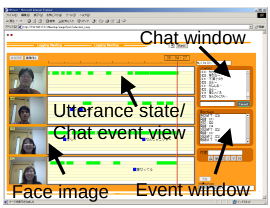

Figure 3. System Windows

Figure 4. Pulse Wave Sensor

Resonance Measurement

To measure the resonance, we calculated the coefficient of correlation between the attendees from all data described above. We calculated the correlation coefficient every second in a 30 second window, which is about twice the period of the 0.07 Hz frequency. Examples of HR data, attendees A and B, are shown in Figure 6 and Figure 7, respectively. Figure 8 shows the coefficient of correlation between A and B. The graph indicates that attendees A and B showed high correlation at around 20 seconds, and low correlation at around 80 seconds.

We then obtained the correlation matrix from the correlation coefficients between the attendees; we call it the "resonance correlation matrix". We estimated the conference state from the resonance correlation matrix. For example, when all 4 attendees are synchronized, the corresponding resonance correlation matrix, **R***a*, is as shown in expression (1).

$$\mathbf{R}a = \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix} \qquad (1)$$

When the 4 attendees were not synchronized at all, **R***b* is as shown in expression (2).

$$\mathbf{R}b = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad (2)$$

As illustrated above, we can use the resonance correlation matrix to assess resonance state.

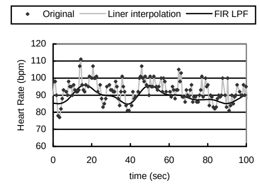

Figure 5. Example of HR Data Processing

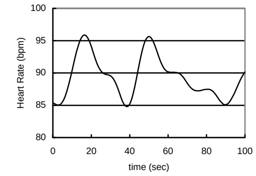

Figure 6. Example of Heart Rate Data (Attendee A)

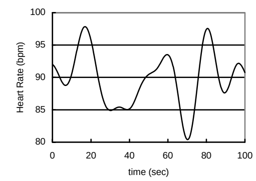

Figure 7. Example of Heart Rate Data (Attendee B)

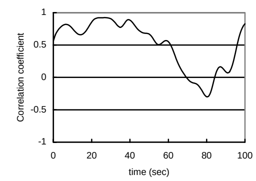

Figure 8. Example of coefficient of correlation between A and B

EXPERIMENT

Conference Recording

A trial conference was recorded [2] as following. The subjects understood the purpose of the experiment; it was for research and the desire was to record the state of a natural conference in detail. As a rehearsal, the subjects used the multimedia conference system to play a few games and hold everyday conversations etc. After being given the topic to be discussed in the conference, the test subjects spent some time thinking about the topic after which the actual recording started. Seven conferences were recorded. An overview of the data obtained is shown in Table 2.

Conference State Analysis

To validate the proposed method, we observed scenes wherein the conference state obviously changed, and analyzed the change based on the resonance assumption. Such changes include changes in the focus of discussion, but we decided to concentrate on the more significant changes that accompany the beginning of or end of participation in the discussion. These changes in participation state include the case where the attendee enters in the middle of a conference, the attendee excuses himself or herself from a conference, and the attendee temporarily dropped out of the conference due to machine trouble etc. In this research, we analyzed the temporary drop out observed in the 1st trial.

During the 1st trial, 4 minutes after starting the conference, ID3 attendee complained about being unable to enter characters into the chat window (Japanese character input feature was disabled due to an operation mistake). She participated in the discussion before this trouble, but after the

Trial	Attendees	Topic	Duration time (min)
1	4 females / telephone operator	survival choice	29
2	4 females / filing clerks	survival choice	37
3	4 females / former architecture students	name choice	61
4	2 males and 2 females / vocational school students	name choice	15
5	1 male and 3 females / band members	name choice	45
6	4 males / university students	new idea	36
7	2 males and 2 females / Chinese	name choice	25

Table 2. Overview of Recorded Data

problem arose, she focused on solving it and failed to participate in the discussion. Figure 9 shows the average of the correlation coefficient obtained from 3 minutes before/after the trouble occurred. 1-3 is the average of correlation coefficient between ID1 and ID3, 1-4 is that of ID1 and ID4, and 3-4 is that of ID3 and ID4. There is no data for ID2 because we failed to measure her pulse wave. As indicated by the correlation between ID1 and ID3 and the correlation between ID3 and ID4, the associations with ID3 weakened while the association between ID1 and ID4, which was independent of ID3, strengthened. This indicates that because of ID3's lack of participation, the association between ID3 and ID1/ID4 via the communication channel weakened and their HRs became asynchronous. Using a similar argument, the discussion between ID1 and ID4 was deepened, their association via the communication channel strengthened, and their HRs started to synchronize.

ID3's trouble was solved about 15 minutes after starting the conference. Figure 10 shows the average of correlation coefficient obtained 3 minutes before/after trouble correction. The 1-4 association, which was independent of ID3, weakened while the 3-4 association strengthened. This indicates that ID3 started to participate in the discussion again because the trouble was solved. The 1-3 association weakened, because the connection used by ID1 to provide advice to ID3, was dropped after the trouble was solved. Regarding the correlation gap of 1-3, 1-4 and 3-4, it is less than the one measured before/after trouble occurrence. The reason is as follows: ID3 didn't participate in the conference just after the trouble, but shortly she started to come back to the conference while solving the trouble. Therefore the connection with ID1 and ID4 was closer than that at the moment of trouble occurrence. This led to the smaller gap.

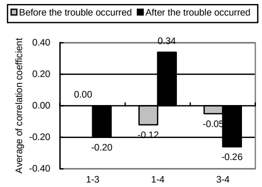

Figure 9. Correlation Change Before/After Trouble Occurred

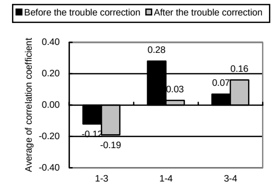

Figure 10. Correlation Change Before/After Trouble Correction

Discussion

Due to ID3's trouble, the associations with ID3 weakened before/after trouble occurrence, while that unrelated to ID3, strengthened. These results show that the correlation coefficient tends to rise while participating in the conference and to fall otherwise. The resonance assumption allows us to explain this characteristic as follows: the association between attendees via the communication channel strengthens while participating in the conference, and weakens otherwise. By utilizing more biosignals and calculating the corresponding resonance correlation matrices, we can better estimate the state of the conference and the closeness/alienation of the attendees, as well as extracting more information about interrelationship of the attendees. This requires the construction of a "resonance state dictionary" which contains the resonance correlation matrix that corresponds to each conference state.

SUMMARY

This paper has discussed a conference state (the discussion is reaching agreement or not, etc.) estimation method that uses only biosignal processing; linguistic understanding is not employed. In conventional dialogue communication research, the physiological characteristic of "entrainment" between the participants has already been reported. We have extended "entrainment" to introduce the concept of "resonance" which we use to grasp the relationship between attendees. We proposed a method that estimates the conference state from measured "resonance" data. First, using a multimedia conference system, we have recorded conference data including time-series records of the participants' heart rate. By observing and analyzing the conference data, we have assessed the "resonance" among the talkers. Using the "resonance correlation matrix", a newly proposed index based on the correlation of the heart rate data, conference participative state could be successfully estimated. In future, we plan to examine in more detail the relationship between the resonance correlation matrix and participation level, for example the change in participation level due to a willful closure of the communication channel. In addition, we will study how to build a "resonance state dictionary" which contains a resonance correlation matrix corresponding to each conference state.

REFERENCES

1. Miyata, A., Fukui, K., Honda, K., Shigeno, H. and Okada, K. A New Way for Analyzing Meeting Using Brain Wave Informations. *IPSJ SIG Tech. Rep.*, 2003-DPS-113 (12), 63-68. (in Japanese)
2. Nakayama, A., Hosoda, M., Indo, T., Kobayashi, M. and Iwaki, S. Video Conference Grasper Based on a Video Conference Corpus and its Multimedia Data Mining. *IPSJ SIG Tech. Rep.*, 2003-GN-49 (20), 115-120. (in Japanese)
3. Sayers, B. M. Analysis of Heart Rate Variability. *Ergonomics*, 16, 1 (1973), 17-32.
4. Watanabe, T., Okubo, M. and Kuroda, T. Analysis of Entrainment in Face-to-Face Interaction Using Heart Rate Variability. *Proc. IEEE RO-MAN'96*, 141-145.
5. Watanabe, T. and Okubo, M. SAKURA: Voice-Driven Embodied Group-Entrained Communication System. *Proc. HCI International 2003*, 2, 558-562.