# Estimation of Empathy Skill Level and Personal Traits Using Gaze Behavior and Dialogue Act During Turn-Changing

Ryo Ishii[1(✉)], Shiro Kumano[2], Ryuichiro Higashinaka[1], Shiro Ozawa[1], and Testuya Kinebuchi[1]

[1] NTT Media Intelligence Laboratories, NTT Corporation, 1-1, Hikari-no-oka, Yokosuka-shi, Kanagawa, Japan
ryo.ishii.ct@hco.ntt.co.jp

[2] NTT Communication Science Laboratories, NTT Corporation, 2-4, Hikaridai, Seika-cho, Atsugi-shi, Kanagawa, Japan

**Abstract.** We explored the gaze behavior towards the end of utterances and dialogue acts (DAs), i.e., verbal-behavior information indicating the intension of an utterance, during turn-keeping/changing to estimate several social skills and personal traits in multi-party discussions. We first collected data on several personal indicators, i.e., Big Five, which measures personal traits, and Davis' Interpersonal Reactivity Index (IRI), which measures empathy skill level, utterances that include DA categories, and gaze behavior, from participants in four-person discussions. We constructed and evaluated models for estimating the scores of these indicators using gaze behavior and DA information. The evaluation results indicate that using both gaze behavior and DAs during turn-keeping/changing is effective for estimating all such scores with high accuracy. It is also possible to estimate these scores with higher accuracy by using the gaze distribution to the current speaker and listener and amount of speaking obtained during the entire discussion. We also found that the IRI scores can be estimated more accurately than those of Big Five.

**Keywords:** Personal traits · Empathy skill level · Gaze behavior · Dialogue act · Turn-taking

## 1 Introduction

Social communication skills are fundamental for successful communication in globalized and multi-cultural societies as they are central to education, work, and daily life. Although there is great interest in the notion of communication skills in scientific and real-life applications, the concept is difficult to generally define due to the complexity of communication, wide variety of related cognitive and social abilities, and huge situational variability [9]. Techniques that involve nonverbal behaviors to estimate communication skills have been receiving much attention.


C. Stephanidis et al. (Eds.): HCII 2021, LNCS 13095, pp. 44–57, 2021.
https://doi.org/10.1007/978-3-030-90963-5_4

For example, researchers have developed models for estimating public speaking skills [38,43], persuasiveness [37], communication skills during job interviews [34] and group work [35], and leadership [42].

Most of these studies used the overall values of verbal/nonverbal behaviors during an entire discussion such as the amount of utterances and physical motion. However, Ishii et al. [20,21] estimated communication skills with high accuracy from such behaviors in a short time during turn-changing. They developed an estimation model of scores of four indexes in Davis' Interpersonal Reactivity Index (IRI) [5] that uses the gaze behavior and dialogue acts (DAs) near the end of utterances during turn-keeping/changing as feature values. The model has a higher estimation accuracy than that using the overall values of verbal/nonverbal behaviors during an entire discussion used in many previous studies on skill estimation. This suggests that behavior during turn-keeping/turn-changing in a very short time is useful for estimating individual empathy skill level.

Ishii et al. [20,21] focused on estimating only the score of one of the four indexes in IRI, which measures empathy skill level. It is necessary to demonstrate whether gaze behavior and DAs are useful for estimating various scores of indicators for measuring the characteristics and social skills of individuals other than IRI.

In this study, we explored whether gaze behavior and DAs during turn-keeping/changing are useful in estimating the scores of the other three IRI indexes, those of Big Five [4], for measuring personal traits, by constructing and evaluating estimation models for them. First, we collected data on the scores from each indicator plus utterances that include the DA categories such as provision, self-discourse, empathy, and turn-yielding, and gaze behavior from participants in four-person discussions. We constructed and evaluated models for estimating theses scores using gaze behavior and DAs. The evaluation results indicate that using both gaze behavior and DA information during turn-keeping/changing is effective for estimating all such scores with high accuracy. It is also possible to estimate these scores with higher accuracy by using the gaze distribution to the current speaker and listener and amount of speaking obtained during the entire discussion. We also found that the IRI scores can be estimated more accurately than those of Big Five.

## 2 Related Work

### 2.1 Personal Traits and Empathy Skill

We describe the importance of Big Five and IRI. Big Five is a taxonomy for five personality traits [4], i.e., "openness to experience (OP)", "conscientiousness (CO)", "extraversion (EX)", "agreeableness (AG)", and "neuroticism (NE)". OP reflects the degree of intellectual curiosity, creativity, and a preference for novelty and variety a person has. It is also describes the extent to which a person is imaginative or independent and depicts a personal preference for a variety of activities over a strict routine. CO reflects the tendency to be organized and dependable, show self-discipline, act dutifully, aim for achievement, and

prefer planned rather than spontaneous behavior. EX reflects energetic, urgency, assertiveness, sociability, the tendency to seek stimulation in the company of others, and talkativeness. AG reflects the tendency to be compassionate and cooperative rather than suspicious and antagonistic towards others. It is also a measure of one's trusting and helpful nature and whether one is generally well-tempered. NE reflects the tendency to be prone to psychological stress, i.e., the tendency to experience unpleasant emotions easily such as anger, anxiety, depression, and vulnerability. It also refers to the degree of emotional stability and impulse control and is sometimes referred to by its low pole, "emotional stability".

As mentioned above, Davis' Interpersonal Reactivity Index (IRI) [5] includes four indexes of empathy: perspective-taking (PT), i.e., the tendency to adopt another's psychological perspective; fantasy (FS), i.e., the tendency to strongly identify with fictitious characters; empathetic concern (EC), i.e., the tendency to experience feelings of warmth, sympathy, and concern toward others; and personal distress (PD), i.e., the tendency to have feelings of discomfort and concern when witnessing others' negative experiences. IRI has been translated into many languages [6,8] and used in a wide variety of fields such as neuroscience [1] and genetics [39]. In this context, some researchers in computer engineering developed models for estimating empathetic statements between people [31,32]. Thus, measuring empathy skill level using IRI is considered invaluable in human communication.

The above three indicators are very important in measuring social skills and personality traits of individuals. Therefore, we focused on estimating the scores of a total of nine indexes of these two indicators. This study is the first attempt to demonstrate the relationships among such scores and gaze behavior and DAs during turn-changing/keeping during discussions.

### 2.2 Verbal and Gaze Behavior During Turn-Changing

Most research on elucidating the mechanism of turn-changing in conversation has been conducted in sociolinguistics. Sacks et al. [41] developed a turn-changing model in which turn-changing can only occur at transition-relevance points near the end of utterances. Several studies have reported that verbal behavior and nonverbal behavior, such as gaze, have an important association with the next speaker and the start of the next utterance [29].

Gaze behavior is especially important for smooth turn-taking. Kendon [29] reported that a speaker gazes at a listener in a two-person conversation as a "turn-yielding cue" at the end of an utterance. The listener glances at the speaker (mutual gazing) then looks away (mutual gaze ends) from the speaker and starts speaking, that is, takes the turn. These findings indicate that the transition of gaze behavior and mutual gaze are important for turn-changing. Some researchers [18,25,27] reported a similar tendency for the speaker to look at the next speaker when yielding the turn in multi-party discussions. On the basis of these findings, many studies have attempted to enable smooth turn-changing using verbal and gaze behavior in human-computer interaction [17,40].

**Fig. 1.** Photograph of multi-party discussion

Several studies have explored the idea of automatically detecting whether turn-changing takes place in multi-party discussions by using gaze behavior near the end of an utterance and other behaviors [3,7,26,27]. In addition to estimating turn-changing, some studies have attempted to estimate who will become the next speaker during turn-changing and when the next utterance will start. Some previous studies have developed estimation models that feature three processing steps to estimate whether turn-changing or turn-keeping will occur, who the next speaker will be during turn-changing, and the start of the next speaker's utterance using the gaze-behavior features of gaze transition patterns (GTPs), which have an n-gram of gaze objects that includes mutual gaze information [12,18,24,25]. GTPs are the most useful patterns found thus far for estimating the next speaker and time of the next utterance.

Thus, people use verbal behavior and nonverbal behavior, such as gaze, near the end of utterances for smooth turn-changing. We assume that a high emotional and empathy skill level is needed to gaze depending on the DA information required for smooth turn-changing. In addition, there is a good possibility that such behavior may differ depending on personality traits. Our key idea is using DA information and gaze behavior during turn-keeping/changing to estimate emotion and empathy skill level and personal traits.

## 3 Corpus Data

In this section, we give details of the corpus of multi-party discussion. The corpus includes eight face-to-face four-person discussions held by four groups of four different people (16 participants in total). In each group, the four participants were Japanese women in their 20's and 30's who had never met before. They sat facing each other (Fig. 1). We labeled the participants, from left to right, P1, P2, P3, and P4. They argued and gave opinions in response to highly divisive questions, such as "Is marriage the same as love?", and needed to reach a conclusion within ten minutes. All four four-person groups took part in two discussions.

**Table 1.** Average values (Avg.) and standard deviations (SD) of 16 people for each indicator's indexes and Pearson's correlation coefficient $r$ between indexes. The yellow and orange boxes are significantly correlated at significance level of $p < .05$ and $p < .01$ as result of uncorrelated test.

| | | Basic statistics | | Correlation coefficient | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | | Big five | | | | | IRI | | | |
| | | Avg. | SD | OP | CO | EX | AG | NE | EC | PT | PD | FS |
| Big five | OP | 4.36 | 0.52 | | 0.081 | 0.515 | 0.110 | −0.456 | 0.071 | −0.107 | 0.072 | −0.034 |
| | CO | 4.01 | 0.87 | 0.081 | | 0.355 | −0.051 | 0.338 | 0.205 | 0.161 | 0.034 | 0.022 |
| | EX | 4.85 | 0.77 | 0.515 | 0.355 | | 0.114 | 0.115 | 0.020 | 0.388 | −0.026 | 0.153 |
| | AG | 5.02 | 0.68 | 0.110 | −0.051 | 0.114 | | −0.615 | −0.006 | −0.026 | −0.206 | 0.105 |
| | NE | 4.34 | 1.06 | −0.456 | 0.338 | 0.115 | −0.615 | | 0.074 | −0.099 | 0.572 | 0.149 |
| IRI | EC | 2.59 | 0.80 | 0.071 | 0.205 | 0.020 | −0.006 | 0.074 | | −0.114 | 0.383 | 0.622 |
| | PT | 3.10 | 0.45 | −0.107 | 0.161 | 0.388 | −0.026 | −0.099 | −0.114 | | −0.094 | −0.101 |
| | PD | 2.52 | 0.54 | 0.072 | 0.034 | −0.026 | −0.206 | 0.572 | 0.383 | −0.094 | | 0.127 |
| | FS | 2.81 | 0.44 | −0.034 | 0.022 | 0.153 | 0.105 | 0.149 | 0.622 | −0.101 | 0.127 | |

The participants' voices were recorded with a pin microphone attached to their chests, and the entire discussions were videoed. Upper body shots of each participant (recorded 30 Hz) were also taken. From the collected data for all eight discussions (80 min in total) and from the recorded data, we constructed a multimodal corpus consisting of the following verbal/nonverbal behaviors and the participants' scores of the indexes of Big Five and IRI.

- Utterances and DAs: We built the utterance unit using the inter-pausal unit (IPU) [30]. The utterance interval was extracted manually from the speech wave. The portion of an utterance followed by 200 ms of silence was used as the unit of one utterance. From the created IPU, backchannels were excluded, and an utterance unit continued from the same person was considered as one utterance turn. IPU pairs adjoined in time, and IPU groups during turn-keeping/changing were created. The data for speech overlaps, i.e., when a listener interrupted during a speaker's utterance or two or more participants spoke simultaneously at turn-changing, were excluded from the IPU pairs for analysis. Eventually, there were 1227 IPUs during turn-keeping and 129 during turn-changing.
- Gaze objects: A skilled annotator manually annotated the gaze objects by using bust/head and overhead views in each video frame. The gaze objects were the four participants (labeled P1, P2, P3, and P4, as mentioned above) and non-persons, i.e., the walls or floor. Three annotators annotated the gaze behavior in our conversation dataset to verify the annotation quality. Conger's Kappa coefficient was 0.887. Based on the benchmarks of a previous study [10], the gaze annotations were of excellent quality.
- Personal traits: All participants were asked to complete a questionnaire that was based on IRI [5] and Big Five [4]. The scores of the five indexes of Big Five, i.e., OP, CO, EX, AG, and NE, and those of the four indexes of IRI, i.e., PT, FS, EC, and PD, for each participant were obtained from their responses. Table 1 shows the average values and standard deviations of 16 people for each

index and Pearson's correlation coefficients $r$ among the 9 indexes of the two indicators. The following significant correlations between indexes were found.

- Correlation of indexes: there is a correlation [between OP and EX and between AG and NE of Big Five as well as between EC and FS of IRI.
- Correlation of indexes between the two indicators: there is a correlation between NE of Big Five and PD of IRI.

Thus, there is a correlation between several indexes.

All verbal and nonverbal behavior data were integrated 30 Hz for visual display using viewer software [36]. This software enables us to annotate multimodal data frame-by-frame and observe the data intuitively.

## 4 Feature Values

We used the gaze behavior and DAs during turn-changing/turn-keeping as feature values for developing estimation models of the scores of each indicator in reference to previous studies [15,20]. In this section, we give details of these feature values.

We first introduce gaze behavior. We focused on GTPs as features of gaze behavior, which are temporal transitions of participant's gaze behavior near the end of utterances according to previous studies [12,18,23,24]. A GTP is expressed as an n-gram, which is defined as a sequence of gaze-direction shifts. We demonstrated that the occurrence frequencies of GTPs differ significantly for a speaker and listener during turn-keeping and a listener who becomes the next speaker (hereafter, called "next-speaker") and listeners who do not become the next speaker (hereafter, called "listeners") during turn-changing. We also demonstrated that a GTP is effective for estimating the next speaker in multiparty discussions. Thus, we used GTPs as gaze-analysis parameters. To generate a GTP, we focused on the gazed object for 1200 ms: 1000 ms before and 200 ms after the utterance since the GTP during 1200 ms is important for turn-taking [12,18,23,24]. A GTP is composed of a person or object classified as "speaker", "listener", or "non-person" and labeled. We considered whether there was mutual gaze and classified gaze behavior using the following seven gaze labels.

- $S$: Person looks at a speaker without mutual gaze (speaker does not look at the listener.).
- $SM$: Person looks at the speaker with mutual gaze (speaker looks at a listener.).
- $L1$, $L2$, $L3$: Person looks at another listener without mutual gaze. Labels $L1$, $L2$, and $L3$ indicate different people. The sitting position does not matter. For example, if P1 who is speaking looks at P2 followed by P3 then P2 again, the gaze transition pattern of P1 is $L1$–$L2$–$L1$.
- $LM1$, $LM2$, $LM3$: Person looks at another listener with mutual gaze. Labels $LM1$, $LM2$, and $LM3$ indicate different people.
- $N$: Person looks at the next speaker without mutual gaze only during turn-changing.

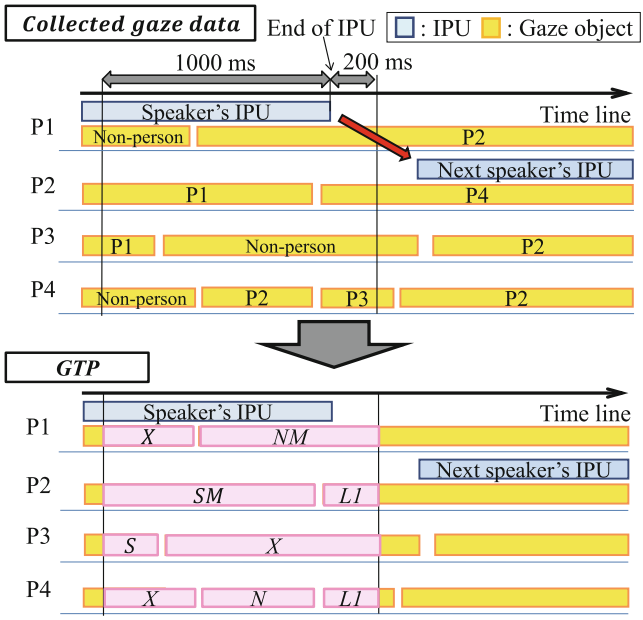


**Fig. 2.** Example of generating GTPs during turn-changing

– $NM$: Person looks at the next speaker with mutual gaze only during turn-changing.
– $X$: Person looks at non-persons, such as the floor or ceiling, i.e., gaze aversion.

Figure 2 shows how GTPs are constructed: P1 finishes speaking, then P2 starts to speak. Person P1 gazes at P2 after she gazes at a non-person during the analysis interval. When P1 looks at P2, P2 looks at P1; that is, there is mutual gaze. Therefore, P1's GTP is $X$-$NM$. Person P2 looks at P4 after making eye contact with P1; thus, P2's GTP is $SM$-$L1$. Person P3 looks at a non-person after looking at P1; thus, P3's GTP is $S$-$X$. Person P4 looks at P2 and P3 after looking at a non-person; thus, P4's GTP is $X$-$N$-$L1$.

A DA for each IPU was extracted using an estimation technique for Japanese [11,33] for DA analysis. This technique can estimate a DA of a sentence from among 33 DA categories using word n-grams, semantic categories (obtained from the Japanese thesaurus Goi-Taikei), and character n-grams. The technique outputs 33 DA categories. We grouped them into the following five major categories.

– Provision: Utterance for providing information
– Self-discourse: Utterance for disclosing oneself
– Empathy: Utterance intending empathy
– Turn-yielding: Utterance intending a listener to speak next (ex. utterance of question, suggestion, or confirmation)
– Others: Utterance not included in the above four categories

About 90% of utterances included the DA categories of Provision, Self-disclosure, Empathy, and Turn-yielding.

Ishii et al. [15,20] demonstrated that the occurrence frequencies of GTPs accompanying each DA category for the speaker and listeners during turn-keeping, and the speaker, next-speaker, and listeners during turn-changing in multi-party discussions, is effective for estimating a participant's EC score. We mainly used them as feature values of the estimation models of the scores of Big Five and IRI in this study.

## 5 Estimation Model

The goal of this study was to demonstrate that the gaze behavior and DAs during turn-keeping/changing are useful for estimating the individuals' scores of all nine indexes of Big Five and IRI. We constructed a model for estimating the index scores of each indicator using GTPs and DA information, one using utterance information such as duration of speaking and number of speaking-turns, and one using simple gaze information (which is the duration of looking at a speaker or listener in a discussion) to compare the usefulness of GTP and DA information. We also constructed two estimation models using GTPs and DA information and using GTPs, DA information, utterance information, and simple gaze information to evaluate the effectiveness of multimodal fusion.

We constructed the estimation models using a SMOreg [28], which implements a support vector machine (SVM) for regression in Weka [2], and evaluated the accuracy of the models and the effectiveness of each feature. The settings of the SVM, i.e., the polynomial kernel, cost parameter ($C$), and hyper parameter of the kernel ($\gamma$), were determined using a grid-search technique. The objective variable is the EC score of each person.

The details of the five estimation models are as follows.

- Chance-level model: This model outputs the mean value of all participants.
- Utterance model: This model uses the ratio of utterances and turns in the discussion.
- Simple-gaze model: This model uses the duration a person was looking at the speaker and listeners in the discussion.
- DA model: This model uses the frequency of occurrence of speech DA category.
- GTP model: This model uses the occurrence frequencies of GTPs. At this time, the occurrence frequencies of GTPs are not classified by DA category.
- GTP+DA model: This model uses the occurrence frequencies of GTPs for each DA category when the person is either the speaker or a listener during turn-keeping and the speaker, next-speaker, or a listener during turn-changing.
- All model: This model uses the ratio of utterances and turns, duration of looking, and occurrence frequencies of GTPs for each DA category. In other words, the features are integrated with an early-fusion method.

**Table 2.** Evaluation results of estimation models. Numbers indicate average absolute errors between estimated and actual scores. Those in parentheses indicate z-score of average absolute errors. Range of scores for each scale, average value obtained from participants, and distribution differ. We use z-score to compare magnitude of error between scales.

| Model | Big five | | | | | IRI | | | |
|---|---|---|---|---|---|---|---|---|---|
| | OP | CO | EX | AG | NE | EC | PT | PD | FS |
| Chance level | 0.422 (0.809) | 0.725 (0.834) | 0.639 (0.828) | 0.516 (0.760) | 0.828 (0.784) | 0.627 (0.783) | 0.306 (0.678) | 0.406 (0.746) | 0.295 (0.676) |
| Simple gaze | 0.693 (1.329) | 0.724 (0.833) | 0.423 (0.549) | 0.383 (0.565) | 0.738 (0.699) | 0.609 (0.761) | 0.232 (0.517) | 0.458 (0.841) | 0.315 (0.724) |
| Utterance | 0.497 (0.953) | 0.883 (1.016) | 0.666 (0.863) | 0.669 (0.986) | 0.970 (0.919) | 0.698 (0.871) | 0.340 (0.754) | 0.472 (0.866) | 0.328 (0.754) |
| GTP | 0.334 (0.640) | 0.520 (0.598) | 0.298 (0.387) | 0.303 (0.447) | 0.437 (0.414) | 0.158 (0.197) | 0.135 (0.299) | 0.041 (0.076) | 0.182 (0.419) |
| DA | 0.431 (0.827) | 0.897 (1.032) | 0.726 (0.942) | 0.656 (0.966) | 1.073 (1.016) | 0.785 (0.980) | 0.447 (0.991) | 0.447 (0.822) | 0.394 (0.905) |
| GTP+DA | 0.288 (0.553) | 0.292 (0.336) | 0.476 (0.618) | 0.482 (0.710) | 0.640 (0.606) | 0.063 (0.079) | 0.040 (0.088) | 0.074 (0.137) | 0.122 (0.280) |
| All | 0.272 (0.521) | 0.458 (0.527) | 0.232 (0.342) | 0.287 (0.423) | 0.222 (0.210) | 0.147 (0.183) | 0.051 (0.114) | 0.061 (0.112) | 0.209 (0.481) |

We used ten-fold cross validation with the data of the 16 participants. The mean absolute error of each estimation model is shown in Table 2.

The simple-gaze model estimated the EX, AG, and NE scores of Big Five and EC and PT of IRI more precisely than the chance-level model. The GTP model estimated all scores more precisely than the chance-level model. The DA model did not estimate any scores more precisely than the chance-level model. Among these models, the GTP model was the most accurate. The GTP+DA model was more accurate for OP, CO of Big Five and EC, PT, and FS of IRI. The accuracy was the highest for CO of Big Five and EC, PT, PD, and FS of IRI for all models.

The All model was the most accurate for OP, EX, AG, and NE of Big Five and PD of IRI. It was found that GTP alone is the most effective feature value in the models that use only one feature value and that the GTP+DA model or All model is most effective. The estimation errors were 0.063 for EC, 0.040 for PT, and 0.061 or less for PD of IRI. Also, among all the indexes, the error was 0.300 or less, and very accurate estimation was possible.

Next, to verify which indexes can be estimated higher and conversely lower, the difference in estimation accuracy among the indexes was analyzed. We compared the errors of the All model' with the highest precision among the other models. The estimation error in each index was divided using the standard deviation of the correct data to obtain a z-score. By comparing the z-scores of the errors in each index, the ease of estimation was examined among the indexes. One-way analysis of variance was conducted on the z-scores of 16 estimation errors of each index, which are the results of a 16 cross-validation, and significant differences were found among the indexes.

**Table 3.** Results of multiple comparison with Fisher's least significant difference method. $*$ indicates that significance level is $p < .05$. $\uparrow$ indicates that estimation error is large compared with comparison target; conversely, $\downarrow$ indicates that estimation error is small compared with comparison target.

| | | Big five | | | | | IRI | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | OP | CO | EX | AG | NE | EC | PT | PD | FS |
| Big five | OP | | n.s. | n.s. | n.s. | $\uparrow *$ | $\uparrow *$ | $\uparrow *$ | $\uparrow *$ | n.s. |
| | CO | n.s. | | n.s. | n.s. | $\uparrow *$ | $\uparrow *$ | $\uparrow *$ | $\uparrow *$ | n.s. |
| | EX | n.s. | n.s. | | n.s. | n.s. | n.s. | $\uparrow *$ | n.s. | n.s. |
| | AG | n.s. | n.s. | n.s. | n.s. | | $\uparrow *$ | $\uparrow *$ | $\uparrow *$ | n.s. |
| | NE | $\downarrow *$ | $\downarrow *$ | n.s. | n.s. | | n.s. | n.s. | n.s. | $\downarrow *$ |
| IRI | EC | $\downarrow *$ | $\downarrow *$ | n.s. | $\downarrow *$ | n.s. | | n.s. | n.s. | $\downarrow *$ |
| | PT | $\downarrow *$ | $\downarrow *$ | $\downarrow *$ | $\downarrow *$ | n.s. | n.s. | | n.s. | $\downarrow *$ |
| | PD | $\downarrow *$ | $\downarrow *$ | n.s. | $\downarrow *$ | n.s. | n.s. | n.s. | | $\downarrow *$ |
| | FS | n.s. | n.s. | n.s. | n.s. | $\uparrow *$ | $\uparrow *$ | $\uparrow *$ | $\uparrow *$ | |

Next, we used multiple comparison and Fisher's least significant difference method to determine which index combination has a difference in estimation error. The results are listed in Table 3. The overall trend can be divided into two groups, i.e., large and small estimation errors, among the indexes. Specifically, errors tended to be smaller for NE of Big Five and EC, PT, and PD of IRI. On the contrary, the errors of OP, CO, EX, and AG of Big Five tended to be large.

## 6 Discussion

Our estimation models using GTP and DA information (i.e. DA+GTP model and All model) have been shown to accurately estimate the index scores of Big Five and IRI. Although gaze behavior and DAs, in the short term, during turn-keeping/changing are carried out unconsciously, they are effective in measuring social skills and personality traits. This is interesting with the findings that have been revealed for the first time. The results also suggest that it is possible to estimate such scores with higher accuracy by using both the gaze distribution to the current speaker and non-speaker obtained from the entire conversation and the amount of utterances in addition to GTPs and DAs.

We also compared the score-estimation accuracy among the nine indexes of the two indicators. NE of Big Five and EC, PT, and PD of IRI tended to have small errors. On the contrary, the errors of OP, CO, EX, and AG of Big Five, and FS of IRI tended to be large. When comparing how many indexes have relatively good accuracy within the indicators, one of the five indexes of Big Five and three of the four indexes of IRI were estimated with high accuracy (i.e. small accuracy error). Overall, Big Five's scores are considered to have large estimation errors and those of IRI have small estimation errors. From these results, the gaze behavior and DAs during turn-keeping/changing are less related to personality

traits and more related to empathy skill level. During turn-changing, it is known that people care about the emotions of others and encourage them to speak using gaze behavior and speech. Therefore, such a result is considered very reasonable.

Finally, we describe the limitations of this research. First, we only used data from 16 people. Also, they were only Japanese women. Therefore, it is necessary to verify how common these results are. However, in spite of a small data set, high-accuracy estimation models could be constructed, so that the gaze behavior and DAs during turn-keeping/changing can be used to estimate the social skills and personality traits of various individuals.

## 7 Conclusion

We examined whether the gaze behavior, specifically GTPs, towards the end of utterance and DA information during turn-keeping/changing are useful for estimating the index scores of various social skills and personal characteristics indicators. It was shown that it is possible to estimate the scores of the nine indexes of Big Five and IRI with high accuracy by using the GTP and DA information during turn-keeping and changing. Although gaze behavior and DAs during turn-keeping/changing is done unconsciously, gaze behavior and DAs are very effective in measuring social skills and personality traits. The results also suggest that it is possible to estimate such scores with higher accuracy by using both the gaze distribution to the current speaker and non-speaker obtained from the entire conversation and the amount of utterances. Furthermore, the IRI scores can be estimated more accurately than those of Big Five. Therefore, the gaze and DAs during turn-keeping/changing may be more related to empathy skill level than personality traits.

For future work, we will explore how effective other behaviors [26], such as head movements [13,16], respiration [12,19,25], and mouth movement [14,22], are during turn-changing for estimating an individual's social skills and personal traits.

## References

1. Banissy, M.J., Kanai, R., Walsh, V., Rees, G.: Inter-individual differences in empathy are reflected in human brain structure. NeuroImage **62**, 2034–2039 (2012)
2. Bouckaert, R.R., et al.: WEKA-experiences with a Java open-source project. J. Mach. Learn. Res. **11**, 2533–2541 (2010)
3. Chen, L., Harper, M.P.: Multimodal floor control shift detection. In: Proceedings of the International Conference on Multimodal Interaction, pp. 15–22 (2009)
4. Costa, P.T., McCrae, R.R.: The NEO personality inventory manual, FL Psychological Assessment Resources (1985)
5. Davis, M.H.: A multidimensional approach to individual differences in empathy **10** (1980)
6. De Corte, K., Buysse, A., Verhofstadt, L.L., Roeyers, H., Ponnet, K., Davis, M.H.: Measuring empathic tendencies: reliability and validity of the Dutch version of the interpersonal reactivity index. Psychologica Belgica **47**, 235–260 (2007)

7. De Kok, I., Heylen, D.: Multimodal end-of-turn prediction in multi-party meetings. In: Proceedings of the International Conference on Multimodal Interaction, pp. 91–98 (2009)
8. Fernandez, A., Dufey, M., Kramp, U.: Testing the psychometric properties of the interpersonal reactivity index (IRI) in Chile: empathy in a different cultural context. Eur. J. Assess. **27**, 179–185 (2011)
9. Greene, J.O., Burleson, B.R.: Handbook of Communication and Social Interaction Skills. Psychology Press, UK (2003)
10. Gwet, K.L.: Handbook of Inter-Rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. Advanced Analytics, LLC (2014)
11. Higashinaka, R., et al.: Towards an open-domain conversational system fully based on natural language processing. In: International Conference on Computational Linguistics, pp. 928–939 (2014)
12. Ishii, R., Kumano, S., Otsuka, K.: Multimodal fusion using respiration and gaze behavior for predicting next speaker in multi-party meetings. In: ICMI, pp. 99–106 (2015)
13. Ishii, R., Kumano, S., Otsuka, K.: Predicting next speaker using head movement in multi-party meetings. In: ICASSP, pp. 2319–2323 (2015)
14. Ishii, R., Kumano, S., Otsuka, K.: Analyzing mouth-opening transition pattern for predicting next speaker in multi-party meetings. In: Proceedings of the International Conference on Acoustics, Speech and Signal Processing, pp. 209–216 (2016)
15. Ishii, R., Kumano, S., Otsuka, K.: Analyzing gaze behavior during turn-taking for estimating empathy skill level. In: Proceedings of the 19th ACM International Conference on Multimodal Interaction, ICMI 2017, pp. 365–373. ACM, New York (2017)
16. Ishii, R., Kumano, S., Otsuka, K.: Prediction of next-utterance timing using head movement in multi-party meetings. In: Proceedings of the 5th International Conference on Human Agent Interaction, HAI 2017, pp. 181–187. ACM, New York (2017)
17. Ishii, R., Miyajima, T., Fujita, K., Nakano, Y.: Avatar's gaze control to facilitate conversational turn-taking in virtual-space multi-user voice chat system. In: Gratch, J., Young, M., Aylett, R., Ballin, D., Olivier, P. (eds.) IVA 2006. LNCS (LNAI), vol. 4133, p. 458. Springer, Heidelberg (2006). https://doi.org/10.1007/11821830_47
18. Ishii, R., Otsuka, K., Kumano, S., Yamamoto, J.: Predicting of who will be the next speaker and when using gaze behavior in multiparty meetings. ACM Trans. Interact. Intell. Syst. **6**(1), 4 (2016)
19. Ishii, R., Otsuka, K., Kumano, S., Yamamoto, J.: Using respiration to predict who will speak next and when in multiparty meetings. ACM Trans. Interact. Intell. Syst. **6**(2), 20 (2016)
20. Ishii, R., Otsuka, K., Kumano, S., Higashinaka, R., Tomita, J.: Analyzing gaze behavior and dialogue act during turn-taking for estimating empathy skill level. In: Proceedings of the 20th ACM International Conference on Multimodal Interaction, ICMI 2018, pp. 31–39. ACM, New York (2018)
21. Ishii, R., Otsuka, K., Kumano, S., Higashinaka, R., Tomita, J.: Estimating interpersonal reactivity scores using gaze behavior and dialogue act during turn-changing. In: Meiselwitz, G. (ed.) HCII 2019, Part II. LNCS, vol. 11579, pp. 45–53. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-21905-5_4
22. Ishii, R., Otsuka, K., Kumano, S., Higashinaka, R., Tomita, J.: Prediction of who will be next speaker and when using mouth-opening pattern in multi-party conversation. Multimodal Technol. Interact. **3**(4), 70 (2019)

23. Ishii, R., Otsuka, K., Kumano, S., Matsuda, M., Yamato, J.: Predicting next speaker and timing from gaze transition patterns in multi-party meetings. In: Proceedings of the International Conference on Multimodal Interaction, pp. 79–86 (2013)
24. Ishii, R., Otsuka, K., Kumano, S., Yamato, J.: Analysis and modeling of next speaking start timing based on gaze behavior in multi-party meetings. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, pp. 694–698 (2014)
25. Ishii, R., Otsuka, K., Kumano, S., Yamato, J.: Analysis of respiration for prediction of who will be next speaker and when? In multi-party meetings. In: Proceedings of the International Conference on Multimodal Interaction, pp. 18–25 (2014)
26. Ishii, R., Ren, X., Muszynski, M., Morency, L.-P.: Can prediction of turn-management willingness improve turn-changing modeling?. In: Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (2020)
27. Jokinen, K., Furukawa, H., Nishida, M., Yamamoto, S.: Gaze and turn-taking behavior in casual conversational interactions. J. TiiS **3**(2), 12 (2013)
28. Keerthi, S.S., Shevade, S.K., Bhattacharyya, C., Murthy, K.R.K.: Improvements to Platt's SMO algorithm for SVM classifier design. Neural Comput. **13**(3), 637–649 (2001)
29. Kendon, A.: Some functions of gaze direction in social interaction. Acta Psychologica **26**, 22–63 (1967)
30. Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., Den, Y.: An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese map task dialogs. Lang. Speech **41**, 295–321 (1998)
31. Kumano, S., Otsuka, K., Matsuda, M., Yamato, J.: Analyzing perceived empathy based on reaction time in behavioral mimicry. IEICE Trans. Inf. Syst. **E97-D**(8), 2008–2020 (2014)
32. Kumano, S., Otsuka, K., Mikami, D., Matsuda, M., Yamato, J.: Analyzing interpersonal empathy via collective impressions. IEEE Trans. Affect. Comput. **6**(4), 324–336 (2015)
33. Meguro, T., Higashinaka, R., Minami, Y., Dohsaka, K.: Controlling listening-oriented dialogue using partially observable Markov decision processes. In: International Conference on Computational Linguistics, pp. 761–769 (2010)
34. Nguyen, L., Frauendorfer, D., Mast, M., Gatica-Perez, D.: Hire me: computational inference of hirability in employment interviews based on nonverbal behavior. IEEE Trans. Multimed. **16**(4), 1018–1031 (2014)
35. Okada, S., et al.: Estimating communication skills using dialogue acts and nonverbal features in multiple discussion datasets. In: Proceedings of the International Conference on Multimodal Interaction, pp. 169–176 (2016)
36. Otsuka, K., Araki, S., Mikami, D., Ishizuka, K., Fujimoto, M., Yamato, J.: Realtime meeting analysis and 3D meeting viewer based on omnidirectional multimodal sensors. In: ACM International Conference on Multimodal Interfaces and Workshop on Machine Learning for Multimodal Interaction, pp. 219–220 (2009)
37. Park, S., Shim, H.S., Chatterjee, M., Sagae, K., Morency, L.-P.: Computational analysis of persuasiveness in social multimedia: a novel dataset and multimodal prediction approach. In: Proceedings of the ACM ICMI, pp. 50–57 (2014)
38. Ramanarayanan, V., Leong, C.W., Feng, G., Chen, L., Suendermann-Oeft, D.: Evaluating speech, face, emotion and body movement time-series features for automated multimodal presentation scoring. In: Proceedings of the ACM ICMI, pp. 23–30 (2015)

39. Rodrigues, S.M., Saslow, L.R., Garcia, N., John, O.P., Keltner, D.: Oxytocin receptor genetic variation relates to empathy and stress reactivity in humans. Proc. Natl. Acad. Sci. U. S. A. **106**, 21437–21441 (2009)
40. Ruhland, K., et al.: A review of eye gaze in virtual agents, social robotics and HCI: behaviour generation, user interaction and perception. Comput. Graph. Forum **34**(6), 299–326 (2015)
41. Sacks, H., Schegloff, E.A., Jefferson, G.: A simplest systematics for the organisation of turn taking for conversation. Language **50**, 696–735 (1974)
42. Sanchez-Cortes, D., Aran, O., Mast, M.S., Gatica-Perez, D.: A nonverbal behavior approach to identify emergent leaders in small groups. IEEE Trans. Multimed. **14**(3), 816–832 (2012)
43. Wortwein, T., Chollet, M., Schauerte, B., Morency, L.-P., Stiefelhagen, R., Scherer, S.: Multimodal public speaking performance assessment. In: Proceedings of the ACM ICMI, pp. 43–50 (2015)