Modeling Turn-Taking Speed and Speaker Characteristics

Kazuyo Onishi
Nara Institute of Science and Technology
RIKEN Guardian Robot Project
onishi.kazuyo.oi5@naist.ac.jp

Hien Onaka
Nara Institute of Science and Technology
RIKEN Guardian Robot Project
onaka.hien.oj5@naist.ac.jp

Koichiro Yoshino
Nara Institute of Science and Technology
RIKEN Guardian Robot Project
Institute of Science Tokyo
koichiro@naist.ac.jp

Abstract

Modeling turn-taking speed while considering speaker characteristics and the relationships between speakers is essential for realizing dialogue systems capable of natural interactions. In this study, we focused on dialogue participants' roles, relationships, and personality, analyzing and modeling turn-taking speeds observed in real conversations. The analysis confirmed that the expression of these attributes—role, relationship, and personality—is closely associated with turn-taking speed. Based on these findings, we constructed a model that predicts the distribution of turn-taking speeds according to each attribute using a gamma distribution. Evaluation results demonstrated that appropriate parameter fitting to the three-parameter gamma distribution enables effective modeling of turn-taking speeds based on participants' roles, relationships, and characteristics.

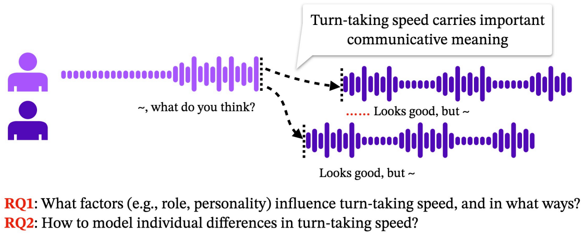


Figure 1: Conceptual Overview of Turn-taking Speed in Conversation

1 Introduction

In conversation, turn-taking refers to the process encompassing both the decision of when a speaker begins or ends an utterance and the interactions involved in that decision. Turn-taking mechanisms are essential for achieving natural dialogue in both human–human and human–dialogue system communication (Skantze, 2021; Hara et al., 2019; Inden et al., 2013; Türker et al., 2017).

Traditional research on turn-taking has provided valuable insights into the factors (e.g., linguistic and non-linguistic cues) that prompt turn transitions and the mechanisms underlying them (Erickson, 1984; Oertel et al., 2012; Jokinen et al., 2010; Cummins, 2012). Building on these findings, research on turn-taking models has progressed significantly; however, discussions have primarily focused on the binary decision of whether to take a turn (Onishi et al., 2024, 2023). Consequently, many dialogue systems implement turn-taking with fixed speeds, lacking the flexibility to adapt to varying conversational contexts and individual speaker characteristics.

In practice, turn-taking speed—the timing of turn exchanges—carries important communicative meaning. For example, a quick response may convey active engagement, while a delayed response can indicate thoughtfulness or reflection. These timing behaviors are employed both consciously and unconsciously in human–human interactions. However, discussions focusing solely on the decision to take a turn overlook these context-sensitive and speaker-dependent adjustments.

Although few studies have focused specifically on turn-taking speed, cross-linguistic and cultural differences have been noted, with Japanese exhibiting a distinctive distribution compared to other languages (Stivers et al., 2009). Furthermore, individual variation and situational factors—such as conversational topic, emotional intensity, and social dynamics—highlight the need for adaptive control of turn-taking speed (Skantze, 2021).

To address this gap, we conducts both empiri-

Proceedings of the 26th Annual Meeting of SIGDIAL, pages 21–31
Aug 25–27, 2025. ©2025 SIGDIAL

cal analysis and predictive modeling of turn-taking speed based on various speaker characteristics. Figure 1 provides a conceptual overview of turn-taking speed in conversation. First, we analyze how turn-taking speed varies with speaker role (expert vs. novice), interpersonal relationship (friend vs. stranger), and Big Five personality traits (openness, agreeableness, conscientiousness, and neuroticism). Building on prior work modeling human's turn taking speed with gamma distributions (Matsuyama et al., 2010), we then develop a predictive model whose mean and variance parameters are conditioned on these speaker characteristics. Experimental results demonstrate that the proposed model more accurately reflects individual differences in turn-taking speed under varied conversational conditions.

Our contributions are summarized as follows:

1. Empirical investigation of how speaker role and interpersonal relationship affect observed turn-taking speed.
2. Demonstration that individual differences in Big Five personality traits significantly influence turn-taking speed.
3. Development of a gamma distribution–based predictive model, tuned to speaker characteristics, for adaptive control of turn-taking speed in dialogue systems.

2 Related Work

In this section, we review previous studies on turn-taking and the influence of individual personality traits on role allocation and speaking behavior in dialogue.

2.1 Previous Research on Turn-Taking

Early experimental studies on turn-taking have demonstrated that the timing for initiating and terminating an utterance is determined by both linguistic and non-linguistic cues (Erickson, 1984; Oertel et al., 2012; Jokinen et al., 2010; Cummins, 2012). In recent years, numerous models addressing the decision-making process for turn transitions have been proposed, with particular attention given to approaches that integrate multimodal information (Onishi et al., 2024, 2023; Fujie et al., 2021; Sakuma et al., 2022; Ward et al., 2018; Lala et al., 2019; Kendrick et al., 2023; Meshorer and Heeman, 2016; Ishii et al., 2022, 2021). Moreover, research that focuses on the speed distribution and adjustment methods of turn-taking—which are influenced by cultural and linguistic differences—has advanced. For instance, Stivers et al. demonstrated that Japanese exhibits a distinctive turn-taking speed distribution compared to other languages (Stivers et al., 2009).

2.2 Relationship Between Personality Traits and Conversational Behavior

It has also been shown that a speaker's role and communication style in dialogue are closely related to their personality traits.

Battistoni and Fronzetti Colladon reported that individuals who play central roles in informal advisory networks tend to exhibit high levels of agreeableness, conscientiousness, and even neuroticism (Battistoni and Colladon, 2014). Licorish and MacDonell revealed that key communicators in global software development teams tend to demonstrate high levels of openness (Licorish and MacDonell, 2015). Moreover, Mehl et al. found that the volume and rhythm of speech in everyday conversation are associated with personality traits such as extraversion, while the updated Big Five model by Soto and John suggests that a detailed breakdown of each factor can capture the multifaceted aspects of individual differences (Mehl et al., 2006; Judge and Ilies, 2002). Furthermore, a meta-analysis by Judge and Ilies comprehensively showed the impact of personality traits on motivation and performance (Soto and John, 2017), findings that are valuable for elucidating the psychological underpinnings of role allocation in dialogue.

Therefore, it is evident that individual personality traits and speaker roles significantly influence conversational dynamics, making it essential to clarify how these factors affect turn-taking behavior.

3 Experimental Design

In this study, we analyze a corpus of human-to-human dialogue to investigate the various factors that influence turn-taking speed. The source code used in the experiments is publicly available[1]. The details of the experiment are described below.

3.1 Corpus

For the purpose of analyzing turn-taking speed, we used the Japanese version of the NoXi

[1] https://github.com/riken-grp/ModelingTurntakingSpeed

Database (Onishi et al., 2024). This corpus was recorded on August 3–4, 2023, on the campus of Nara Institute of Science and Technology (NAIST) by the authors. Audio and video data were recorded synchronously; the audio, video, annotations[2], recording program[3], and analysis program[4] are publicly available for research purposes.

For the recordings in this corpus, participants were instructed to take on the roles of either an "expert" or a "novice." They then engaged in conversations on specific topics such as travel, soccer, and research, communicating via monitors with near-zero delay. The conversation topics were selected from areas of mutual interest, and the dialogues lasted between 10 and 30 minutes, comprising a total of 22 dialogues (approximately 6.8 hours) by 20 participants. In addition, questionnaire responses evaluating the participants' personality traits were collected. In this study, we analyze the relationships between turn-taking speed and speaker roles, interpersonal relationships, and individual personality traits.

3.2 Speaker Roles

In this study, the two speaker roles in the dialogue are designated as expert and novice. The expert, who is knowledgeable about the topic, is responsible for providing content related to the subject. In contrast, the novice, who lacks expertise on the topic, responds to the expert's input and asks simple clarifying questions. It is expected that speaker roles have a significant impact on turn-taking speed. Experts, who require additional time to process information, are predicted to have slower turn-taking speeds, whereas novices, who provide smoother responses, are assumed to have faster turn-taking speeds.

3.3 Interpersonal Relationships

The interpersonal relationship between speakers is believed to have a significant influence on their conversational style and turn-taking speed. In this study, participants are categorized as either friends or first-time acquaintances, and the characteristics of turn-taking speed are compared between these groups. We hypothesize that conversations among friends, which typically occur in a relaxed atmosphere, would tend to exhibit slower turn-taking speeds. Conversely, conversations between first-time acquaintances, which due to tension involves a stronger desire to avoid delayed responses, are likely to exhibit faster turn-taking speeds.

3.4 Personality Traits (BIG5)

In this study, participants' personality traits were assessed using the Japanese version of the Ten Item Personality Inventory (TIPI-J) (Oshio et al., 2012, 2014), which comprises five trait scales each measured by one positively keyed and one negatively keyed item. The TIPI-J is a concise diagnostic tool that evaluates the Big Five personality dimensions through ten questionnaire items. The Big Five traits are described as follows:

- **Extraversion**: A personality trait characterized by liveliness, sociability, and positive affect;
- **Openness**: A personality trait that reflects intellectual curiosity and an interest in new experiences;
- **Agreeableness**: A personality trait that reflects compassion and cooperativeness toward others;
- **Conscientiousness**: A personality trait characterized by responsibility, planning, and self-discipline;
- **Neuroticism**: A personality trait that indicates sensitivity to stress and anxiety.

In this study, for each trait, participants with a positive-item score of 5 or higher and a negative-item score of 3 or lower were classified into the "high" group, those with both positive- and negative-item scores of 4 were classified into the "middle" group, and those with a positive-item score of 3 or lower and a negative-item score of 5 or higher were classified into the "low" group; analyses were conducted primarily using the high and low groups.

These personality traits are expected to play important roles in conversation. For example, individuals with high agreeableness may be more likely to adjust their turn-taking to match their conversational partner. Clarifying the relationship between these personality traits and turn-taking speed could contribute to the development of dialogue systems that foster high levels of trust with users.

[2]For inquiries: onishi.kazuyo.oi5@naist.ac.jp
[3]https://github.com/ahclab/NoXiRecorder
[4]https://github.com/ahclab/NoXiAnalysis

3.5 Modeling for Gamma Distribution

In this study, we observed that the distribution of turn-taking speeds is right-skewed and asymmetric. We approximated this empirical distribution using a three-parameter gamma distribution, whose shape parameter k, scale parameter θ, and location parameter μ allow flexible control over skewness and variance. The gamma function $\Gamma(k)$ generalizes the factorial to the continuous domain, and its probability density function is

$$f(x; k, \theta, \mu) = \frac{(x-\mu)^{k-1} e^{-(x-\mu)/\theta}}{\theta^k \, \Gamma(k)}, \quad x > \mu.$$

The flexibility of the gamma distribution in adjusting tail heaviness and peak sharpness makes it particularly well-suited to capture the observed characteristics of our turn-taking speed data (Matsuyama et al., 2010). By allowing the location parameter μ to assume a negative value, we can incorporate negative floor-transfer offsets (i.e. overlapping speech) into the support of the gamma distribution while preserving its skewness and tail-control properties. Hence, in this study we permit $\mu < 0$.

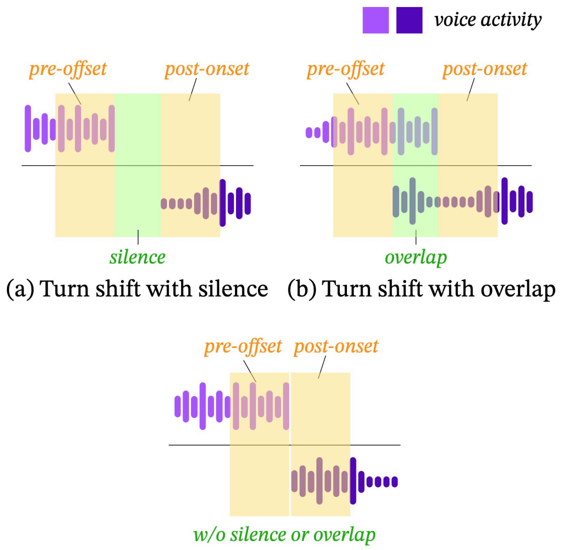
To tailor the gamma distribution to specific subgroups, we recalculated k and θ based on the subgroup's sample mean $\mathbb{E}[X]$ and variance $\mathrm{Var}(X)$ as follows:

$$\theta = \frac{\mathrm{Var}(X)}{\mathbb{E}[X]-\mu}, \quad k = \frac{\left(\mathbb{E}[X]-\mu\right)^2}{\mathrm{Var}(X)}, \quad \mu = \mu_0.$$

This makes it possible to reproduce the turn-taking speed distribution from the mean and variance of the turn-taking speed in the domain to be reproduced. This method has the potential to be used as a model for spoken dialogue systems, and is essential for realizing domain-specific spoken dialogue systems.

3.6 Definition of Turn-Taking

In this study, turn-taking is detected from speech segments using a rule-based approach. This method, which relies solely on speech segments rather than manual annotations, is well-suited for analyzing large amounts of data. We classify transitions into three types—silence, overlap, and direct handover without silence or overlap. Note that prior work has defined only transitions involving silence; in this study, we additionally define overlap and direct handover cases. Figure 2 illustrates these three cases. We set both *pre-offset* and *post-onset* to 0.3 seconds.



(a) Turn shift with silence (b) Turn shift with overlap

(c) Turn shift without silence or overlap

Figure 2: Definition of Turn-Taking

1. **Classification based on the characteristics of the interval $[t_1, t_2]$:**
   - **(a) Turn transition with silence**: For all time points in $[t_1, t_2]$, neither speaker exhibits any speech activity (i.e., silence).
   - **(b) Turn transition with overlap**: For all time points in $[t_1, t_2]$, both User 1 and User 2 are simultaneously active (i.e., overlap).
   - **(c) Turn transition without silence or overlap**: In $[t_1, t_2]$, one speaker stops exactly when the other starts, with no silent gap and no overlap.

2. **Conditions immediately preceding t_1 (*pre-offset*):**
   - User 1 is active.
   - User 2 is inactive.

3. **Conditions immediately following t_2 (*post-onset*):**
   - User 1 is inactive.
   - User 2 is active.

4. An interval that satisfies the above conditions is defined as a turn-taking.

It should be noted that our definition requires the first speaker to become inactive following the transition. This condition ensures that we are measuring a complete transfer of the speaking floor, and

Table 1: Turn-Taking Speed for the Entire Corpus

Metric	Overall
Samples (N)	3148
Mean (s)	0.349
Variance (s^2)	0.513
Overlap (%)	29.9

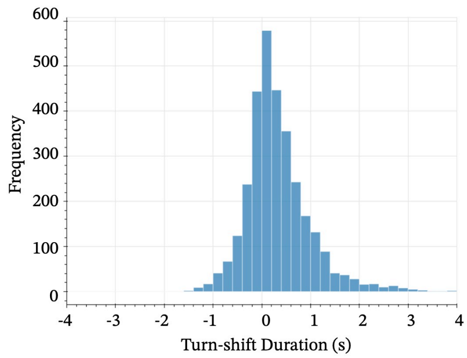


Figure 3: Distribution of Turn-Taking Speed Across the Entire Corpus

consequently, it excludes instances of backchanneling, where a listener provides a short utterance while the primary speaker continues to hold the floor.

4 Results

4.1 Turn-taking Speed Analysis

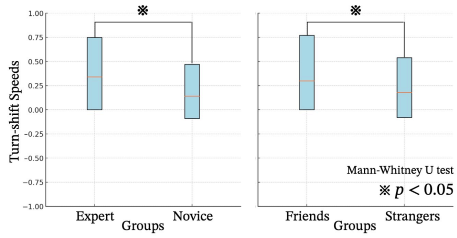
4.1.1 Corpus-wide Turn-Taking Speed

Figure 3 shows the distribution of turn-taking speed across the entire corpus, and Table 1 summarizes the overall statistics. Compared with previous studies that reported a turn-taking speed of 7 milliseconds for Japanese (Stivers et al., 2009), a slightly slower value was observed; however, the overall shape of the turn-taking speed distribution is nearly identical to that reported in earlier research (Godfrey et al., 1992).

4.1.2 Turn-Taking Speed by Speaker Role and Relationship

Table 2 and Figure 4 present the statistical information on turn-taking speed based on speaker roles (novice to expert, expert to novice) and speaker relationships (friends versus first-time acquaintances).

The mean turn-taking speed from novice to expert is 0.455 seconds, while that from expert to novice is 0.247 seconds, with a significant difference confirmed by the Mann–Whitney U test ($p < 0.05$). In addition, the mean turn-taking speed



(a) By Role (b) By Relationship

Figure 4: Turn-Taking Speed by Speaker Role and Relationship

Table 2: Turn-Taking Speed by Speaker Role and Relationship (*$p < 0.05$)

Group	Smp. (N)	Mean (s)	Var. (s^2)	Overlap (%)
Role*				
Novice to Expert	1543	0.455	0.614	25.7
Expert to Novice	1605	0.247	0.395	34.0
Relationship*				
Friends	1391	0.462	0.637	25.7
Strangers	1757	0.259	0.397	33.2

among friends is 0.462 seconds, compared to 0.259 seconds for first-time acquaintances, and this difference was also found to be significant ($p < 0.05$).

Figure 4 shows box plots of the turn-taking speed distributions for each group. The results indicate that longer turn-taking speeds tend to occur in turns from novice to expert and in conversations between friends.

4.1.3 Relationship between BIG5 and Turn-Taking Speed

Table 3 presents the statistical information on turn-taking speed based on the BIG5 personality traits (Extraversion, Openness, Agreeableness, Conscientiousness, and Neuroticism). Using the Mann–Whitney U test, we found significant differences ($p < 0.05$) in Openness, Agreeableness, Conscientiousness, and Neuroticism. Although the mean values for Openness did not differ greatly, the high-Openness group exhibited a larger variance. For Agreeableness, the high-Agreeableness group demonstrated faster turn-taking speeds, while the low-Agreeableness group showed a much larger variance. Regarding Conscientiousness, the high-Conscientiousness group had faster turn-taking speeds. Similarly, the high-Neuroticism group also exhibited faster turn-taking speeds.

Table 3: Relationship between BIG5 Personality Traits and Turn-Taking Speed ($*p < 0.05$)

Group	Smp. (N)	Mean (s)	Var. (s^2)	Overlap (%)
Extraversion				
High	1360	0.430	0.681	29.2
Low	1512	0.324	0.387	27.9
Openness*				
High	2470	0.325	0.534	31.2
Low	306	0.362	0.285	25.2
Agreeableness*				
High	1762	0.313	0.487	31.6
Low	424	0.572	0.990	26.4
Conscientiousness*				
High	727	0.233	0.431	34.8
Low	1636	0.400	0.465	26.1
Neuroticism*				
High	1593	0.257	0.382	32.4
Low	1555	0.443	0.630	27.4

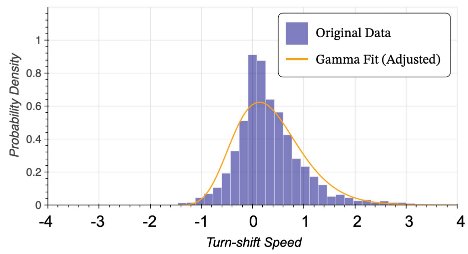


Figure 5: Gamma Distribution with Initial Parameters (k_0, θ_0, μ_0)

4.2 Modeling Turn-Taking Speed

4.2.1 Initial Parameter Estimation

We fit a three-parameter gamma distribution to the entire dataset using maximum likelihood estimation in SciPy. The estimated parameters were:

$$k_0 = 9.8458, \quad \theta_0 = 0.2131, \quad \mu_0 = -1.7495.$$

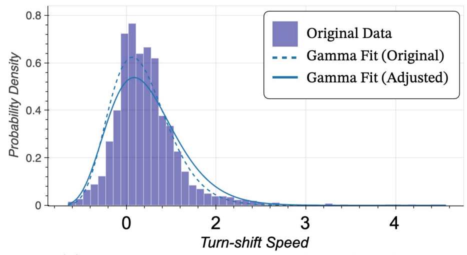
Figure 5 shows the empirical histogram (purple bars) overlaid with the fitted gamma density (solid blue curve), with the vertical axis denoting probability density. The fit closely matches the data, exhibiting only minor deviations near the mode. We assess goodness-of-fit using the Kolmogorov–Smirnov (KS) D-statistic and the Hellinger similarity (percentage), obtaining $D = 0.0612$ and a Hellinger similarity of 99.00%.

4.2.2 Parameter Adjustment via Mean and Variance

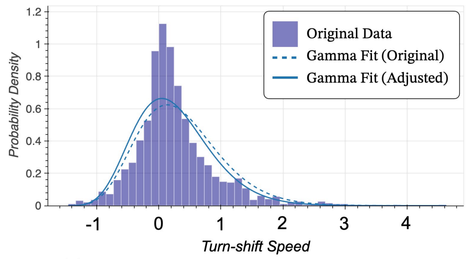
We adjust the parameters based on group-specific means and variances to capture domain-specific characteristics. Table 4 presents the recalculated k and θ for the items introduced in Section 4.1.

Table 4: Domain-Specific Parameters

Groups	k	θ
Expert	7.9144	0.2786
Novice	10.0939	0.1978
Friend	7.6733	0.2882
Stranger	10.1704	0.1975
Extraversion (High)	6.9701	0.3127
Extraversion (Low)	11.1270	0.1864
Openness (High)	8.0590	0.2574
Openness (Low)	15.6275	0.1351
Agreeableness (High)	8.7380	0.2361
Agreeableness (Low)	5.4457	0.4263
Conscientiousness (High)	9.1157	0.2175
Conscientiousness (Low)	9.9254	0.2165
Neuroticism (High)	10.5275	0.1906
Neuroticism (Low)	7.6345	0.2872



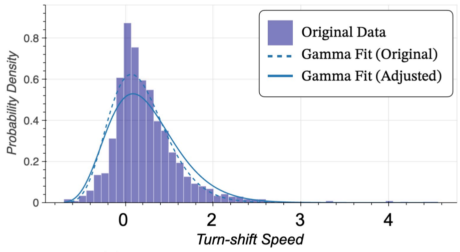
(a) Novice to Expert turn-taking distribution



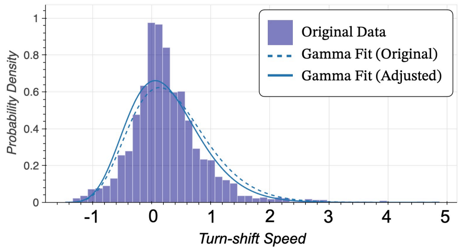
(b) Expert to Novice turn-taking distribution

Figure 6: Gamma Distribution Fits After Mean/Variance Adjustment for Expert and Novice Groups

Figure 6 compares the original (full-data fit) and adjusted (subgroup-specific) gamma distributions for the expert and novice groups, while Figure 7 does the same for the friend and stranger groups. In these figures, the purple bars denote the subgroup histogram, the solid blue line indicates the subgroup-specific gamma density, and the dashed blue line shows the full-data fit. As these figures illustrate, the adjusted distributions (solid lines) more closely follow the shape of the actual data histograms (purple bars) than the original, unadjusted distribution (dashed line). For instance, in Figure 7, the adjusted distribution for the 'friend' group is visibly shifted to the right, graphically confirming our earlier finding that friends exhibit a slower turn-taking pace. The gamma distribu-

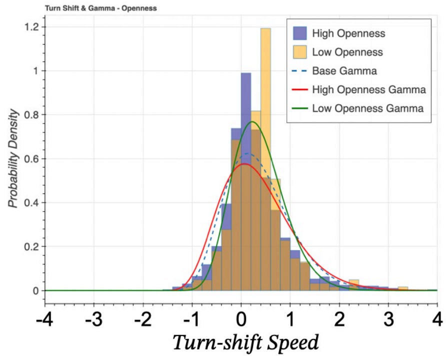


(a) Friend turn-taking distribution

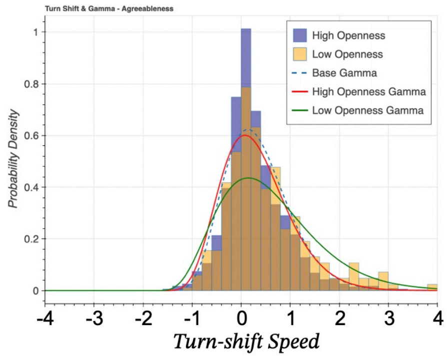


(b) Stranger turn-taking distribution

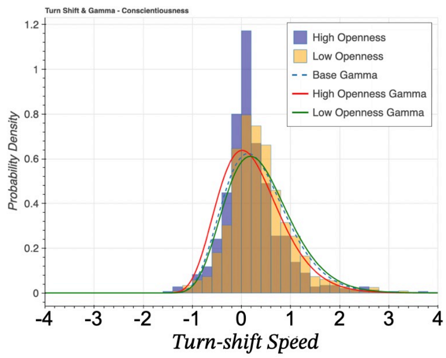
Figure 7: Gamma Distribution Fits After Mean/Variance Adjustment for Friend and Stranger Groups



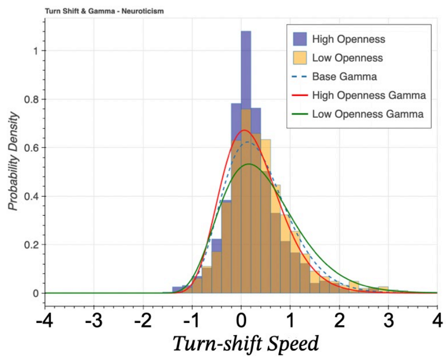
(a) Openness turn-taking distribution



(b) Agreeableness turn-taking distribution



(c) Conscientiousness turn-taking distribution



(d) Neuroticism turn-taking distribution

Figure 8: Gamma Distribution Fits After Mean/Variance Adjustment for BIG5 Groups

tions for each individual trait are provided in the appendix.

Figure 8 presents the gamma distributions for high (purple) and low (yellow) BIG5 trait groups, with probability density on the vertical axis and turn-taking time on the horizontal axis. The dotted line indicates the base distribution, while the red and green curves show the adjusted distributions for the high and low groups, respectively.

Table 5 summarizes goodness-of-fit metrics for each distribution. The reduced (D)-statistic after adjustment indicates a closer match to the empirical data, demonstrating that mean-variance tuning effectively captures domain-specific turn-taking patterns. Hellinger similarity improved scores in many cases, showing accuracies ranging from 98.44

5 Discussion

The results of this study suggest that turn-taking speed varies significantly depending on the roles, relationships, and personality of the dialogue participants. This indicates that the timing of utterances in dialogue is strongly influenced not only by linguistic factors but also by psychological and social factors.

5.1 Speaker Roles

The observation that the turn-taking speed from novice to expert is slower than that from expert to novice can be attributed to experts leading the discussion while novices are required to provide more thoughtful responses. This may be due to an environment in which delayed responses are accept-

Table 5: Evaluation Metrics for Gamma Distribution Fits

Groups	KS D-statistic	Hellinger similarity (%)
Full data vs. Expert	0.0960	98.95
Expert adjusted	0.0826	98.71
Full data vs. Novice	0.1421	98.43
Novice adjusted	0.0807	98.77
Full data vs. Friend	0.0954	99.05
Friend adjusted	0.0892	98.88
Full data vs. Stranger	0.1106	98.59
Stranger adjusted	0.0752	98.91
Full data vs. Openness (High)	0.0894	98.72
Openness (High) adjusted	0.0933	98.44
Full data vs. Openness (Low)	0.1155	98.03
Openness (Low) adjusted	0.0907	99.20
Full data vs. Agreeableness (High)	0.0879	98.94
Agreeableness (High) adjusted	0.0845	98.83
Full data vs. Agreeableness (Low)	0.0807	98.07
Agreeableness (Low) adjusted	0.0897	99.10
Full data vs. Conscientiousness (High)	0.1506	98.15
Conscientiousness (High) adjusted	0.0981	98.51
Full data vs. Conscientiousness (Low)	0.0927	99.13
Conscientiousness (Low) adjusted	0.0689	99.20
Full data vs. Neuroticism (High)	0.1283	98.39
Neuroticism (High) adjusted	0.0859	98.75
Full data vs. Neuroticism (Low)	0.0788	99.17
Neuroticism (Low) adjusted	0.0748	99.02

able because the accuracy and persuasiveness of the information provided by experts are prioritized. Another possibility is that this latency stems from the experts themselves, who may need more time to consider how to best phrase their next utterance for the novice.

5.2 Interpersonal Relationships

Our results showed a slower turn-taking pace among friends. This phenomenon can be explained by the differing social dynamics. In close friendships, there is greater silence tolerance, where pauses are comfortable and allow for thoughtful responses. Additionally, their conversations often involve more complex or personal topics, naturally requiring more time to formulate a reply. In contrast, first-time acquaintances likely experience a stronger pressure for impression management, leading them to respond more quickly to avoid awkward gaps and signal engagement.

5.3 Influence of Personality Traits

The impact of individual personality differences on turn-taking speed is also intriguing. For instance, individuals with low agreeableness scores exhibited slower turn-taking speeds and a much larger variance, which may reflect a tendency to speak in a self-centered manner without attempting alignment with the conversational partner. Additionally, individuals with high conscientiousness scores demonstrated rapid responses, possibly due to their strong sense of responsibility and planning abilities. On the other hand, the finding that individuals with high neuroticism scores show faster turn-taking speeds suggests that emotional instability may serve as a factor that accelerates response speed during dialogue. It is notable that high scores on four of the five traits were associated with faster turn-taking. The lack of a significant effect for Extraversion in our data could suggest this trait influences other conversational aspects, such as utterance length, more than timing, though a larger dataset may be needed to confirm this.

5.4 Gamma Distribution Approximation

By fitting a gamma distribution to the observed turn-taking speeds and then adjusting its parameters using the sample mean and variance, we showed that a simple probabilistic model can closely reproduce the empirical distributions for both experts and novices. While the gamma distribution is a reasonable choice, we acknowledge that the fit is not perfect, as the empirical distributions appear to have fatter tails. Future work could explore alternative distributions that might better capture these characteristics. The differing parameter sets highlight how domain-specific timing profiles can be generated. This approach points toward a mechanism for dialogue systems to adapt their turn-taking behavior dynamically—modulating response delays according to user role or conversational context.

Overall, our findings underscore the multifaceted nature of turn-taking timing and lay the groundwork for future work on adaptive, personality- and context-aware dialogue systems.

6 Limitations and Future Work

Despite the novel insights provided by this study, several limitations should be noted, and directions for future work are identified.

6.1 Data Scale Constraints

First, the JaNoXi corpus comprises only 22 sessions from 20 participants (6.8 hours), which may constrain the generalizability of our conclusions. Our analysis was limited to Japanese dialogues, so cross-linguistic and cultural validity remain untested. We have yet to measure the practical benefits of adaptive speed control in live dialogue systems with end users.

6.2 Statistical Testing Scope

While multiple comparison corrections (e.g., Bonferroni or Holm procedures) are important for controlling the family-wise error rate when conducting multiple tests, their application would introduce substantial complexity beyond the focused, one-factor analysis presented here. Therefore, we deliberately limited our investigation to independent effect analyses without applying correction methods. Future work that examines multidimensional interactions among factors is encouraged to incorporate appropriate multiple comparison adjustments to ensure statistical rigor.

7 Conclusion

In this study, we investigated how turn-taking speed varies in human–human dialogue and demonstrated its dependence on speaker role (expert vs. novice), interpersonal relationship (friends vs. strangers), and Big Five personality traits. Our empirical analysis revealed clear associations between these factors and both the shape and scale parameters of turn-taking intervals.

Building on these findings, we developed a three-parameter gamma distribution–based predictive model whose parameters are conditioned on subgroup mean and variance. This simple adjustment method reduced the Kolmogorov–Smirnov statistic and achieved Hellinger similarities at max 99.20%, confirming its high fidelity in reproducing observed speed distributions.

Overall, our approach lays the groundwork for adaptive turn-taking strategies in dialogue systems. By dynamically modulating response timing according to user characteristics, future conversational agents and human–robot interfaces can deliver more natural, context-aware interactions.

8 Acknowledgements

This work was supported in part by KAKENHI Grant Number 23K24910. This work was conducted under the RIKEN Graduate Student Research Associate Program.

A Appendix

Figure 9 shows the distribution of turn-taking speed for each of the BIG5 personality trait groups using box plots.

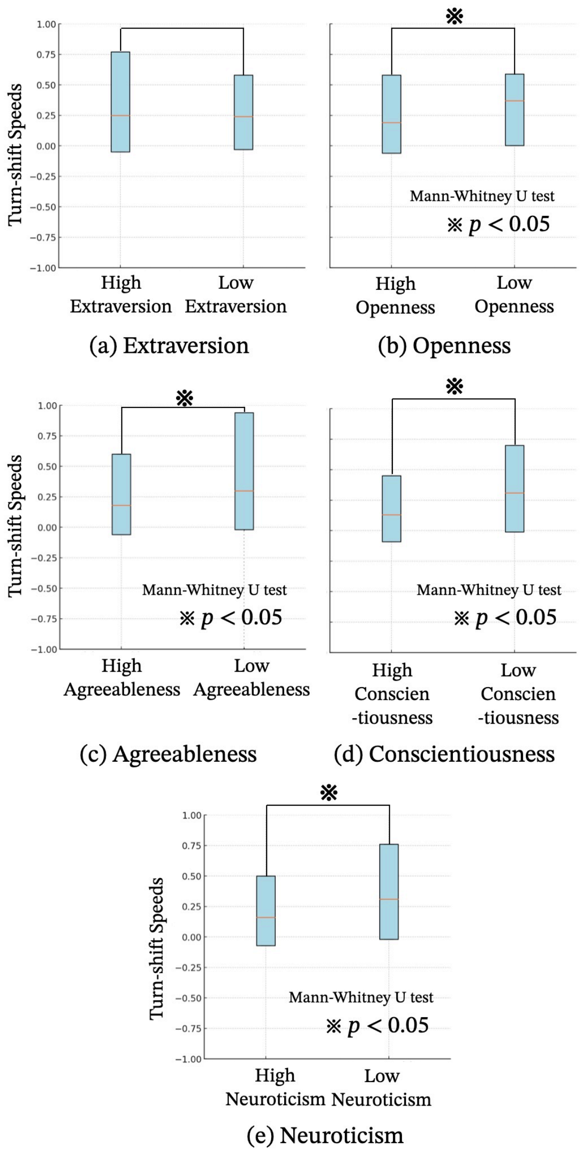


Figure 9: Distribution of Turn-Taking Speed by BIG5 Personality Traits

References

Elisa Battistoni and Andrea Fronzetti Colladon. 2014. Personality correlates of key roles in informal advice networks. *Learning and Individual Differences*, 34:63–69.

Fred Cummins. 2012. Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals. *Language and Cognitive Processes*, 27(10):1525–1549.

Frederick D Erickson. 1984. Conversational organization: Interaction between speakers and hearers.

Shinya Fujie, Hayato Katayama, Jin Sakuma, and Tetsunori Kobayashi. 2021. Timing generating networks: Neural network based precise turn-taking timing prediction in multiparty conversation. In *22nd Annual Conference of the International Speech Communication Association, INTERSPEECH 2021*, pages 3771–3775. International Speech Communication Association.

John J Godfrey, Edward C Holliman, and Jane McDaniel. 1992. Switchboard: Telephone speech corpus for research and development. In *Acoustics, speech, and signal processing, ieee international conference on*, pages 517–520. IEEE Computer Society.

Kohei Hara, Koji Inoue, Katsuya Takanashi, and Tatsuya Kawahara. 2019. Turn-taking prediction based on detection of transition relevance place. In *INTERSPEECH*, pages 4170–4174.

Benjamin Inden, Zofia Malisz, Petra Wagner, and Ipke Wachsmuth. 2013. Timing and entrainment of multimodal backchanneling behavior for an embodied conversational agent. In *Proceedings of the 15th ACM on International conference on multimodal interaction*, pages 181–188.

Ryo Ishii, Xutong Ren, Michal Muszynski, and Louis-Philippe Morency. 2021. Multimodal and multitask approach to listener's backchannel prediction: Can prediction of turn-changing and turn-management willingness improve backchannel modeling? In *Proceedings of the 21st ACM International Conference on Intelligent Virtual Agents*, pages 131–138.

Ryo Ishii, Xutong Ren, Michal Muszynski, and Louis-Philippe Morency. 2022. Trimodal prediction of speaking and listening willingness to help improve turn-changing modeling. *Frontiers in Psychology*, 13:774547.

Kristiina Jokinen, Masafumi Nishida, and Seiichi Yamamoto. 2010. On eye-gaze and turn-taking. In *Proceedings of the 2010 workshop on eye gaze in intelligent human machine interaction*, pages 118–123.

Timothy A Judge and Remus Ilies. 2002. Relationship of personality to performance motivation: a meta-analytic review. *Journal of applied psychology*, 87(4):797.

Kobin H Kendrick, Judith Holler, and Stephen C Levinson. 2023. Turn-taking in human face-to-face interaction is multimodal: gaze direction and manual gestures aid the coordination of turn transitions. *Philosophical Transactions of the Royal Society B*, 378(1875):20210473.

Divesh Lala, Koji Inoue, and Tatsuya Kawahara. 2019. Smooth turn-taking by a robot using an online continuous model to generate turn-taking cues. In *2019 International Conference on Multimodal Interaction*, pages 226–234.

Sherlock A Licorish and Stephen G MacDonell. 2015. Communication and personality profiles of global software developers. *Information and Software Technology*, 64:113–131.

Kyoko Matsuyama, Kazunori Komatani, Ryu Takeda, Toru Takahashi, Tetsuya Ogata, and Hiroshi G Okuno. 2010. Analyzing user utterances in barge-in-able spoken dialogue system for improving identification accuracy. In *INTERSPEECH*, pages 3050–3053.

Matthias R Mehl, Samuel D Gosling, and James W Pennebaker. 2006. Personality in its natural habitat: manifestations and implicit folk theories of personality in daily life. *Journal of personality and social psychology*, 90(5):862.

Tomer Meshorer and Peter A. Heeman. 2016. Using past speaker behavior to better predict turn transitions. In *Interspeech*.

Catharine Oertel, Marcin Włodarczak, Jens Edlund, Petra Wagner, and Joakim Gustafson. 2012. Gaze patterns in turn-taking. In *Thirteenth annual conference of the international speech communication association*.

Kazuyo Onishi, Hiroki Tanaka, and Satoshi Nakamura. 2023. Multimodal voice activity prediction: Turn-taking events detection in expert-novice conversation. In *Proceedings of the 11th International Conference on Human-Agent Interaction*, pages 13–21.

Kazuyo Onishi, Hiroki Tanaka, and Satoshi Nakamura. 2024. Multimodal voice activity projection for turn-taking and effects on speaker adaptation. *IEICE Transactions on Information and Systems*, advpub:2024HCP0002.

Atsushi Oshio, Shingo Abe, Pino Cutrone, and Samuel D Gosling. 2014. Further validity of the japanese version of the ten item personality inventory (tipi-j). *Journal of Individual Differences*.

Atsushi Oshio, ABE Shingo, and Pino Cutrone. 2012. Development, reliability, and validity of the japanese version of ten item personality inventory (tipi-j). *Japanese Journal of Personality/Pasonariti Kenkyu*, 21(1).

Jin Sakuma, Shinya Fujie, and Tetsunori Kobayashi. 2022. Response Timing Estimation for Spoken Dialog System using Dialog Act Estimation. In *Proc. Interspeech 2022*, pages 4486–4490.

Gabriel Skantze. 2021. Turn-taking in conversational systems and human-robot interaction: a review. *Computer Speech & Language*, 67:101178.

Christopher J Soto and Oliver P John. 2017. The next big five inventory (bfi-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power. *Journal of personality and social psychology*, 113(1):117.

Tanya Stivers, Nicholas J Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter De Ruiter, Kyung-Eun Yoon, and 1 others. 2009. Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, 106(26):10587–10592.

Bekir Berker Türker, Zana Buçinca, Engin Erzin, Yücel Yemez, and T Metin Sezgin. 2017. Analysis of engagement and user experience with a laughter responsive social robot. In *Interspeech*, pages 844–848.

Nigel G Ward, Diego Aguirre, Gerardo Cervantes, and Olac Fuentes. 2018. Turn-taking predictions across languages and genres using an lstm recurrent neural network. In *2018 IEEE Spoken Language Technology Workshop (SLT)*, pages 831–837. IEEE.