Virtual Character-based Study of the Combined Effect of Turn-taking Behavior and Speech Speed on Conversational Atmosphere

Masahide Yuasa

Shonan Institute of Technology, Fujisawa, Kanagawa 251-8511, Japan
yuasa@sc.shonan-it.ac.jp

Abstract. In the field of human-computer interaction, researchers have explored methods to enhance the turn-taking abilities of conversational agents and robots in interactions with humans. Previous studies have shown that variations in turn-taking patterns (e.g., overlaps and gaps) can influence the perceived conversational atmosphere, and thus, their effect on human perceptions has been examined. However, the combination impact of turn-taking behavior and speech speed on conversational atmosphere remains underexplored. To address this, we developed a conversational simulator featuring virtual conversational characters with simple shapes and meaningless utterances to investigate how the combination of turn-taking and speech speed influences the identification of a conversational atmosphere. The characters followed basic turn-taking rules, and the simulator controlled their turn-taking behaviors. We conducted an experiment in which participants observed scenes generated by the simulator. A two-way repeated-measures ANOVA was performed, examining the effects of turn-taking behaviors (overlap, no-gap-no-overlap, gap) and speech speed (fast, medium, slow). Results indicated that fast speed was perceived as creating a competitive atmosphere, even when turn-taking adhered to no-gap-no-overlap patterns. This finding will contribute to developing conversational agents/robots that have the same capacity to judge conversational atmosphere as humans.

Keywords: Conversational Agent, Character, Turn-taking, Speech Speed, Overlap, Gap.

1 Introduction

In the area of human-computer interaction, it is well established that turn-taking behaviors provide not only turn management but also influence the impressions formed during everyday conversations [1]. Consequently, how turn-taking patterns affect human senses has been studied [1,2]. Reference [3] examined the relationship between conversational atmospheres and turn-taking patterns, suggesting that individuals might infer conversational atmospheres based on the degree of overlap or gaps during interactions.

However, the combined effect of speech speed and turn-taking patterns on the identification of conversational atmospheres has not been thoroughly investigated. Speech

speed is believed to also influence the perception of conversation, and the interaction between speech speed and turn-taking may significantly impact how individuals perceive conversations. Despite this, the specific effects of these variables remain largely unknown.

In this study, we developed a conversational simulator featuring virtual characters to examine how the combination of speech speed and turn-taking behavior influences the identification of conversational atmospheres. The findings from this study could contribute to the development of conversational agents and robots that can perceive conversational atmospheres, such as whether a conversation feels friendly or competitive, as illustrated in Fig. 1. Moreover, this research may aid in designing conversational agents that can evaluate whether the atmosphere of a human conversation is appropriate to participate in.

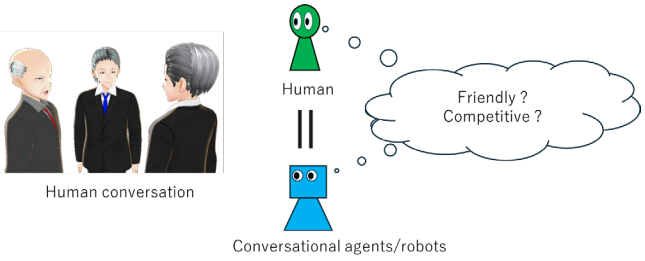

Fig. 1. Contributions of this study. Our investigation contributes to the development of conversational agents/robots that can identify conversational atmospheres like humans.

2 Previous Studies

Previous studies have reported that turn-taking behaviors, such as overlaps and gaps, not only manage the transition of speaking turns among participants [4-6] but also convey social meanings [7], influence the tension level of conversation [8], and reflect interpersonal attitudes [9].

Ter Maat et al. investigated the impressions created by different turn-taking behaviors. Their study examined the emotional responses and impressions associated with turn-taking patterns [1,2], identifying factors such as agreeableness and assertiveness through variations in turn-taking behaviors. However, their research was limited to vocal expressions, as they did not incorporate the use of agents' body movements.

Another study [3] proposed a classification of conversational types based on the degree of overlap or gap between speaking turns. The first type, referred to as a "competitive conversation," was characterized by excessive overlaps [8,10]. The second type, termed 'friendly conversation,' involved moderate overlap [10]. The third type, known as "formal/well-mannered conversation," was distinguished by the presence of gaps between turns [11,12]. Although these conversational types are categorized based on turn-taking behaviors, the relationship between turn-taking patterns and speech speed has not been thoroughly established.

It has also been reported that speech speed varies based on factors such as politeness, gender, and whether the speaker is a native or non-native speaker [13-16]. Yu et al.

found that the speaking rate of a conversational agent affected the perception of the speaker's trustworthiness [17]. Dowding et al. investigated users' preferences for system speech rate, finding that fast speakers preferred faster system speech, while slower speakers preferred slower speech [18]. Xie et al. examined the influence of varying speech rates in conversational agents. They reported that users preferred the feedback speed to increase in response to their own faster speech rate [19].

While studies have explored the effects of changes in speech speed, the combined effect of speech speed and turn-taking behaviors on the overall conversational atmosphere remains largely unexplored. If conversational agents and robots have only knowledge about the effects of one element and lack knowledge regarding the effect of the combined effect of two factors, the difference between agents/robots and humans will occur. Thus, our study aimed to gain insights into the combined effect of these two factors, integrate our knowledge for conversational behaviors, and contribute to the development of conversational agents and robots that can perceive conversational atmospheres like humans do.

3 Conversational Simulator and Turn-taking Rules

Building on previous studies [3,8], we developed a conversational simulator featuring virtual characters. Fig. 2 illustrates a conversation scene generated by the simulator. The simulator was built using WebGL, and the 3D characters were sourced from an external website [20]. To minimize biases related to character appearance, the virtual characters were designed as abstract, simple shapes with no facial expressions. Their heads and bodies could rotate, allowing them to look at the speaking character by aligning their head and body movements. The characters' mouth movements were synchronized with their vocal utterances.

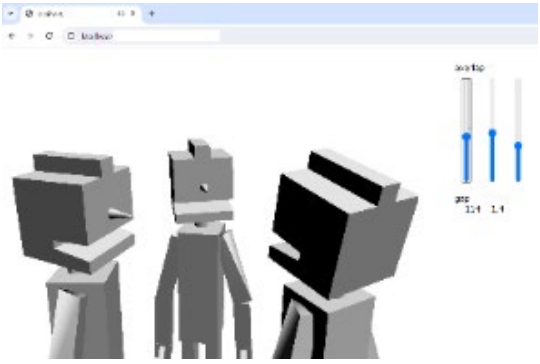

Fig. 2. A conversation scene in our simulator

To focus solely on conversational behavior, the virtual characters spoke in meaningless utterances such as "blah-blah-blah, yadda-yadda-yadda, or banana-banana-banana." This approach, inspired by previous experiments that used non-human or filtered voices [1-3], helped isolate the effects of conversational behavior from verbal content.

Each utterance lasted for approximately 6.6 s, with consistent length across characters. Drawing on the findings of earlier research [3], we implemented typical conversational behaviors for the virtual characters (see Fig. 3):

1. When a character begins speaking, the other characters direct their gaze toward the speaker.
2. While speaking, the active character looks straight ahead.
3. Before finishing, the speaking character turns to look at the next designated speaker.
4. Once the speaker finishes, the next speaker starts; in cases of overlap, the new speaker does not wait for the previous one to complete their turn.

The simulated conversation then returns to the first behavior, with the next speaker selected automatically before the current speaker finishes, and this cycle continues.

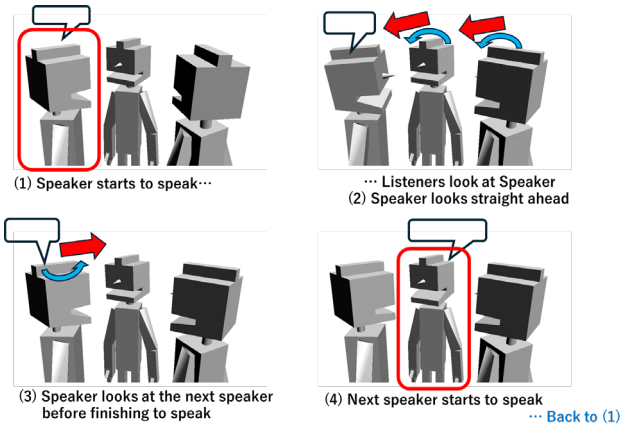

Fig. 3. Character behaviors in our simulator

4 Experimental Design

4.1 Turn-taking patterns and speeds of speech

We prepared three types of turn-taking patterns: overlap, no-gap-no-overlap, and gap, based on the experimental results from a previous study [3] and our preliminary experiment. The duration of the overlap was set to approximately -1.9 s, which might be perceived as sufficient overlap. Similarly, the gap duration was set to approximately +1.9 s, which might be understood as a sufficient pause[1].

We utilized a voice synthesis tool (TTSMP3 [21]) to generate meaningless vocalizations (e.g., "blah-blah-blah"). The speech speed of these utterances was set to approximately 110 words per minute (wpm), which is generally perceived as a clear speaking

[1] These durations were based on formulated 'k-values' in the previous study [3]: k = 70 (-1.9 s), k = 100 (0 s), and k = 130 (+1.9 s).

rate [22]. This speed served as the baseline for our experiment. Using the audio processing library SoundTouchJS [23], we adjusted the speech speed of the utterances to create two additional conditions: approximately 150 wpm (1.4x faster than the baseline), representing fast speech, and approximately 90 wpm (0.8x slower than the baseline), representing slow speech [24].

The pitch of the utterance remained unchanged when the speech speed was adjusted. Although the length of the utterances was stretched or compressed based on the speech speed, the number of words remained consistent across all conditions.

4.2 Experimental Procedure and Participants

Using the three turn-taking patterns and three speech speeds, we produced nine unique scenes with the conversational simulator and recorded them as nine separate videos for participant evaluation. The resolution of each video was 450 x 275 pixels, which was sufficient for participants to answer the questions. Each video lasted approximately 30–40 s. Participants were not required to watch the videos to the end. The videos were presented to participants in random order.

While watching each video, participants were asked to rate it on a 7-point Likert scale. For example, they could rate the friendliness of the conversation as "not friendly" (-3)), "neither" (0) to "friendly" (+3). Based on previous studies [3] and our preliminary experiments, we selected three items, “competitive,” “friendly,” and “well-mannered” to evaluate each video.

Prior to viewing the videos, participants were informed that the study focused on modeling human conversations. They were also told that the behaviors of the virtual characters in the videos were modeled on human conversational behaviors and voices observed in real interactions. Participants were asked to infer how the three characters were communicating with each other based solely on their movements and vocalizations.

The participants were recruited for the experiment from Amazon Mechanical Turk (a crowdsourcing website). An informed consent procedure approved by our institution was used. Only those who consented participated in the experiment. Participants were paid points that could be redeemed on shopping sites.

4.3 Hypothesis

Based on previous findings, we hypothesized that speech speed would influence participants' comprehension of conversational atmospheres. The following hypothesis was formulated:

Hypothesis: Participants’ ratings would be influenced not only by turn-taking patterns but also by speech speed, and an interaction effect between turn-taking and speech speed would be observed.

5 Results

The experiment included 56 participants after excluding inappropriate responses.[2] Fig. 4 shows the average ratings for "competitive," "friendly," and "well-mannered." (The values were converted from "-3 to +3" to "0 to 6.") A two-way repeated-measures analysis of variance (ANOVA) was conducted on the responses, Factor A being turn-taking patterns and Factor B being speech speed.

There was a significant interaction effect between turn-taking and speech speed for "competitive" ($F(4,220) = 2.509$, $p = 0.042 < .05$, $\eta_p^2 = 0.044$). Table I shows the results of the simple main effects of "competitive" in each condition. In the multiple comparisons using the Bonferroni method, there were significant differences. Table II shows the results of the comparisons. We can see that the B1(Fast) level differed from other levels and that the "No-gap-no-overlap" pattern had no difference from the "Overlap" pattern. Similarly, the A1(Overlap) level had no significant difference in speech speed.

Regarding "friendly," there were significant main effects for Factor A($F(2,110) = 46.876$, $p < .01$, $\eta_p^2 = 0.46$) and Factor B($F(2,110) = 7.141$, $p < .01$, $\eta_p^2 = 0.115$), and there was no interaction effect. In the multiple comparisons using the Bonferroni method, there were significant differences. For Factor A, "Overlap" < "No-gap-no-overlap" and "Overlap" < "Gap" ($p < .05$, mean square error (MSE) = 3.447, alpha' = 0.0167). Regarding Factor B, "Fast" < "Medium" and "Fast" < "Slow" ($p < .05$, MSE = 2.861, alpha' = 0.0167).

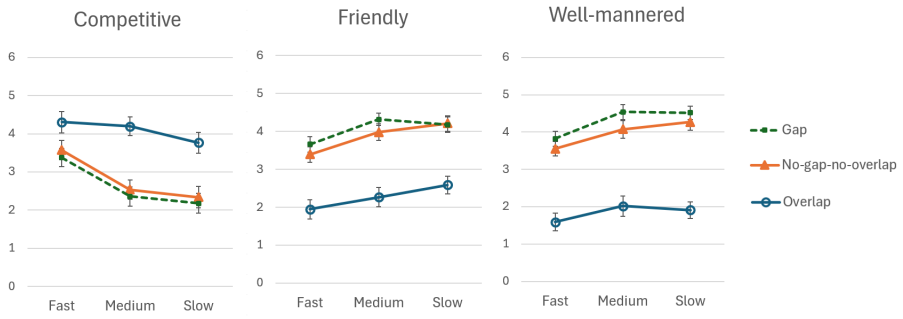

Fig. 4. Averages ± S.E. for "competitive," "friendly," and "well-mannered."

[2] We used participants with a specified qualification of 'HIT Approval Rate greater than 90%', which meant a high grade on the site. Participant-age groups were 21–30 (13%), 31–40 (41%), 41–50 (23%), 51–60 (14%), and 61 and above (9%)

Table 1. Results of simple main effects for “competitive.” (+ $p < .10$, * $p < .05$, ** $p < .01$)

A: Turn-taking		B: Speech Speed	
A at B1(Fast)	F(2,110) = 4.54 *	B at A1(Overlap)	F(2,110) = 2.79 +
A at B2(Medium)	F(2,110) = 14.60 **	B at A2(No-gap-no-overlap)	F(2,110) = 9.41 **
A at B3(Slow)	F(2,110) = 11.31 **	B at A3(Gap)	F(2,110) = 10.40 **

Table 2. Multiple comparisons using the Bonferroni method for “competitive.” (* $p < .05$, alpha'= 0.0167)

A at B1(Fast) Level (MSE = 2.9562)	B at A1(Overlap) Level (MSE = 1.6152)
Overlap = No-gap-no-overlap n.s.	Fast = Medium n.s.
Overlap > Gap *	Fast = Slow n.s.
No-gap-no-overlap = Gap n.s.	Medium = Slow n.s.

A at B2(Medium) Level (MSE = 3.9464)	B at A2(No-gap-no-overlap) Level (MSE = 2.6083)
Overlap > No-gap-no-overlap *	Fast > Medium *
Overlap > Gap *	Fast > Slow *
No-gap-no-overlap = Gap n.s.	Medium = Slow n.s.

A at B3(Slow) Level (MSE = 3.7904)	B at A3(Gap) Level (MSE = 2.2425)
Overlap > No-gap-no-overlap *	Fast > Medium *
Overlap > Gap *	Fast > Slow *
No-gap-no-overlap = Gap n.s.	Medium = Slow n.s.

In the case of “well-mannered,” there were significant main effects for Factor A($F(2,110) = 67.613$, $p < .01$, $\eta_p^2 = 0.551$) and Factor B($F(2,110) = 9.317$, $p < .01$, $\eta_p^2 = 0.145$), and there was no interaction effect. In the multiple comparisons using the Bonferroni method, there were significant differences. Regarding Factor A, “Overlap” < “No-gap-no-overlap” and “Overlap” < “Gap” ($p < .05$, MSE = 4.405, alpha' = 0.0167). Regarding Factor B, “Fast” < “Medium” and “Fast” < “Slow” ($p < .05$, MSE = 1.925, alpha' = 0.0167).

6 Discussion

6.1 Verification of hypothesis and considerations from experimental results

The experimental results showed a significant interaction effect between turn-taking patterns and speech speed for the "competitive" rating. At the fast speech level (B1), the "no-gap-no-overlap" pattern did not differ significantly from the "overlap" pattern, even though the "overlap" pattern was perceived as more "competitive" than the "no-gap-no-overlap" pattern at other speech speeds. This result indicates that participants' responses were influenced not only by turn-taking patterns but also by speech speed. It

suggests that speech speed plays a significant role in the perception of conversational atmospheres, partially supporting our hypothesis.

However, for the "friendly" and "well-mannered" ratings, there was no interaction effect between turn-taking and speech speed. This suggests that the turn-taking pattern operated independently of speech speed, with both factors affecting the ratings of "friendly" and "well-mannered" separately. Moreover, in the case of the "no-gap-no-overlap" pattern, fast speech did not reduce the ratings of "friendly" and "well-mannered," even though fast speech significantly increased the ratings of "competitive."

It is possible that the reason fast speech influenced the perception of "competitive" turn-taking is related to the role of turn-taking in managing the control of speaking turns at the end of utterance. Reference [25] examined the behavior of speakers toward the end of their utterances and found that certain behaviors express a strong desire to take control of the conversation. Fast speech may create the impression of a strong desire to take the floor, which could explain the increase in "competitive" ratings. Further experiments focusing on competitive turn-taking are needed to explore these relationships in greater detail.

6.2 Limitations

To isolate the effects of turn-taking and speech speed, and to avoid content-related biases, we used meaningless words in the experiment. However, future studies should employ more practical utterances to better reflect real-world conversations. Additionally, while we used simplified characters to eliminate bias related to appearance, future experiments with more realistic characters are needed to obtain deeper insights into how the conversational atmosphere is perceived.

Furthermore, cultural differences in conversational norms must be considered. In our experiment, 55% of participants were from the U.S., 43% from India, and 2% from other countries. Although this study is the first to explore the combined effects of turn-taking and speech speed, further experiments with a more diverse range of participants are required to gain more comprehensive insights.

7 Conclusions

In this study, we conducted an experiment to investigate the relationship between turn-taking behaviors and speech speed. The results confirmed an interaction effect, showing that fast speech speed significantly influences the perception of competitive conversations. This finding can contribute to the development of conversational agents and robots capable of making judgments about the conversational atmosphere (e.g., determining whether a conversation is competitive) and joining human conversations. Additionally, our use of abstract characters and meaningless words in a bottom-up approach provides insights into cognitive psychology, which aims to the understanding of the internal models.

Further experiments addressing generational gaps, cultural differences, and other factors are necessary to deepen these findings.

Conflicts of Interest

The author declares no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Acknowledgment

We are grateful to the persons who participated in our experiments.

References

1. Ter Maat, M. and Heylen, D.: Turn Management or Impression Management? Proceedings of IVA2009, 5773, Springer, https://doi.org/10.1007/978-3-642-04380-2_51 (2009).
2. Ter Maat, M., Truong, K. P. and Heylen, D.: How Turn-Taking Strategies Influence Users' Impressions of an Agent. Proceedings of IVA2010, vol. 13, pp. 441–453, https://doi.org/10.1007/978-3-642-15892-6_48 (2010).
3. Yuasa, M.: Investigation of the Relationship between Turn-taking Behaviors and Conversational Atmospheres using Virtual Characters, The Transactions of Human Interface Society, 2024, Vol. 26, Issue 2, 259-262, https://doi.org/10.11184/his.26.2_259 (2024).
4. Sacks, H., Schegloff, E.A., and Jefferson, G.: A Simplest Systematics for the Organisation of Turn-Taking for Conversation. Language, 50(4), 696–735, http://dx.doi.org/10.2307/412243 (1974).
5. Goffman, E.: Interaction Ritual: Essays on face-to-face behavior. New York, NY: Doubleday Anchor (1967).
6. Schegloff, E.A.: Overlapping talk and the organization of turn-taking for conversation. Language in Society, 29, 1-63, https://doi.org/10.1017/S0047404500001019 (2000).
7. Coates, J.: No gap, lots of overlap: Turn-taking patterns in the talk of women friends. In Researching Language and Literacy in Social Context, Graddol, David, Maybin, Janet, and Stierer, Barry (eds.). Clevedon: Multilingual Matters, 177–192 (1994).
8. Yuasa, M.: Can Animated Agents Help Us Create Better Conversational Moods? Proceedings of HCII2014 (2014).
9. Ravenet, B., Cafaro, A., Biancardi, B., Ochs, M., and Pelachaud, C.: Conversational Behavior Reflecting Interpersonal Attitudes in Small Group Interactions. Proceedings of IVA2015, pp. 375–388, https://doi.org/10.1007/978-3-319-21996-7_41 (2015).
10. Dunne, M., and Ng, S. H.: Simultaneous Speech in Small Group Conversation: All-Together-Now and One-at-a-Time? Journal of Language and Social Psychology, 13, 45–71, https://doi.org/10.1177/0261927X94131004 (1994).
11. Fairclough, N.: Discourse and Social Change. Polity Press (1992).
12. Hakulinen, A.: Conversation types. In Handbook of Pragmatics, Jef Verschueren, Jan-Ola Östman, Jan Blommaert and Chris Bulcaen(eds.), pp. 101-120. John Benjamins (1999).
13. Quené, H.: Multilevel modeling of between-speaker and within-speaker variation in spontaneous speech tempo, J. Acoust. Soc. Am., 123(2), 1104–1113, https://doi.org/10.1121/1.2821762 (2008).
14. Jacewicz, E., Fox, RA., and Wei, L.: Between-speaker and within-speaker variation in speech tempo of American English, J Acoust Soc Am., 128(2), 839–850, https://doi.org/10.1121/1.3459842 (2010).

15. Munro, M. J., and Derwing, T. M.: Modeling perceptions of the accentedness and comprehensibility of L2 speech: The Role of Speaking Rate. Studies in Second Language Acquisition, 23(4), 451–468, http://www.jstor.org/stable/44486957 (2001).
16. Ofuka, E., McKeown, J., and Waterman, M., and Roach, P.: Prosodic cues for rated politeness in Japanese speech, Speech Commun., 32, 199–217, https://api.semanticscholar.org/CorpusID:10838259 (2000).
17. Yu, Y., and Levitan, S.I.: What makes a conversational agent sound trustworthy? Exploring the role of acoustic-prosodic factors. Proc. Speech Prosody 2024, pp. 1240–1244, https://doi.org/10.21437/SpeechProsody.2024-250 (2024).
18. Dowding, S., Gutwin, C., and Cockburn, A.: User speech rates and preferences for system speech rates, International Journal of Human-Computer Studies, 184, https://doi.org/10.1016/j.ijhcs.2024.103222 (2024).
19. Xie, Y., Qu, J., Zhang, Y. et al.: Speaking, fast or slow: how conversational agents' rate of speech influences user experience, Univ Access Inf Soc, 2023, https://doi.org/10.1007/s10209-023-01000-2 (2023).
20. Social-exp.site, https://social-exp.site/, last accessed 2024/9/17.
21. TTSMP3, https://ttsmp3.com/, last accessed 2024/9/17.
22. What's your speech rate?, https://www.write-out-loud.com/speech-rate.html, last accessed 2024/9/17.
23. SoundTouchJS, https://github.com/cutterbl/SoundTouchJS/, last accessed 2024/9/17.
24. How fast do you speak and type?, https://www.typingmaster.com/speech-speed-test/, last accessed 2024/9/17.
25. Yuasa, M., Mukawa, N., Kimura, K., Tokunaga, H., and Terai, H.: An Utterance Attitude Model in Human-Agent Communication: From Good Turn-taking to Better Human-Agent Understanding, CHI Extended Abstracts 2010, pp. 3919-3924, (2010).