話者特性に基づくターンテイキング速度の分析

大西一誉 [1,2] 大中緋慧 [1,2] 吉野幸一郎 [1,2,3]
[1] 奈良先端科学技術大学院大学
[2] 理化学研究所ガーディアンロボットプロジェクト
[3] 東京科学大学
{onishi.kazuyo.oi5,onaka.hien.oj5,koichiro}@naist.ac.jp

概要

本研究では，対話時のターンテイキングにおけるターン速度の違いとその要因を明らかにするため，話者の役割，関係性，性格特性に着目し分析を行った．その結果，ターン速度は話者の役割や関係性，個人特性に大きく依存することが確認された．初学者と専門家間の対話では，初学者が発話の準備に時間を要するためターン速度が遅くなる傾向が確認された．また，友人同士の会話では沈黙が許容されやすく，初対面に比べてターン速度が遅い結果が得られた．さらに，BIG5 の性格特性において，開放性，協調性，勤勉性および神経症傾向がターン速度に影響を与えることが明らかになった．

1 はじめに

会話において発話者がどのタイミングで発言を開始あるいは終了するかの決定と，決定に関わる種々のインタラクションを包含してターンテイキングと呼ぶ．人間同士の円滑なコミュニケーションや人間-対話システムにおいて，自然な会話を実現するためにターンテイキングの技術は必要不可欠である [1, 2, 3, 4].

従来のターンテイキング研究では，ターンが切り替わる因子（言語特徴/非言語特徴）やそのメカニズムについて多くの知見が得られてきた [5, 6, 7, 8]. これらを元にターンテイキングモデルの研究が活発に行われてきたが，これまでのところターンを取るか否かの決定に関する議論に収束している [9, 10, 11, 12, 13, 14, 15, 16, 17, 18]．しかししターンテイキングにおいてはターンを取るか否かだけでなく，ターンを交替を起こすタイミングに関わるターン速度に関する議論は十分に行われていない．そのため，多くの対話システムにおけるターンテイキング手法は，ターン速度が固定されているものも多く，発話状況や個人特性に応じた柔軟な調整が十分に実現されていない．

このように，ターン速度に着目した先行研究は少ないが，言語や文化の違いがターン速度に影響を与えることは指摘されており，日本語は他の言語と比較して独特なターン速度の分布を持つことが明らかになっている [19]．また，ターン速度は非常に個人差が大きいことも容易に想像ができ，発話するドメインや状況などにも大きく左右される可能性が高い [1].

本研究では，話者の役割，話者同士の関係性，個人特性に着目して，それらがターン速度に与える影響を分析し明らかにすることを目的とする．これらはターンテイキングにおける人間らしいタイミング調整を実現するための基礎となるだけでなく，対話システムの自然性や応答品質の向上に寄与するものである．

2 実験デザイン

本研究では，人同士の対話コーパスを分析の対象とし，その中でターン速度に与える様々な因子について調査する．以下に実験の詳細を述べる．

2.1 コーパス

本研究では，ターン速度の分析を目的として日本語版 NoXi Database を利用した [9]．本コーパスは 2023 年 8 月 3 日から 4 日にかけて，著者が奈良先端科学技術大学院大学（NAIST）の校内で収録したものである．音声および動画データを同期的に記録し，音声・動画・注釈[1)]，収録プログラム[2)]および分析プログラム[3)]は，研究利用に限り公開されている．

本コーパスで参加者は専門家と初学者の役割に分

1） 要お問い合せ: onishi.kazuyo.oi5@naist.ac.jp
2） https://github.com/ahclab/NoXiRecorder
3） https://github.com/ahclab/NoXiAnalysis

かれ，旅行，サッカー，研究といった特定のトピックについて，限りなく遅延が 0 に近い状態のモニター越しに対話を行っている．会話のトピックは，双方が興味を持つ分野から選定され，対話の長さは 10 分から 30 分程度，合計 22 対話（約 6.8 時間分），20 名の参加者から構成されている．また，参加者の個人特性を評価する質問紙の回答が付随する．今回はこのコーパスを用い，話者の役割，話者の関係性，個人特性それぞれとターン速度の関係について分析する．

2.2 話者の役割

対話における話者の役割を専門家と初学者に分類して分析を行う．専門家はそのトピックに精通した役割を担い，その分野についての話題を提供する必要がある．初学者は対象トピックについて知識が乏しい役割を担い，専門家の話題に対しての応答や，簡単な疑問点の質問などを行う．話者の役割はターン速度に大きく影響すると考えられ，専門家が情報を処理する時間を必要とし，ターン速度が遅くなる一方で，初学者はスムーズな応答を行うためターン速度が速くなると予想される．

2.3 話者の関係性

話者同士の関係性は，対話のスタイルやターン速度に大きな影響を与える可能性があると考えられる．本研究では，参加者を友人同士と初対面に分類し，それぞれのターン速度の特徴を比較した．友人同士の会話では，リラックスした雰囲気での対話が行われるため，ターン速度が遅くなる傾向があると予測される．一方で，初対面の参加者同士の会話では，緊張感から応答の遅延を避けるため，ターン速度が速くなる可能性が高い．

2.4 個人特性 (BIG5)

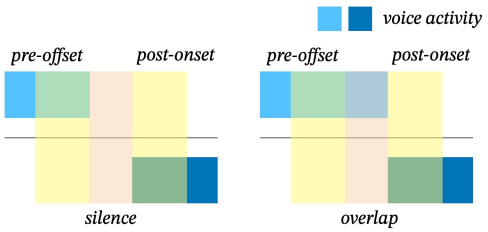
本研究では，参加者の個人特性を測るために TIPI-J（Ten Item Personality Inventory - Japanese）を用いる [20, 21]．TIPI-J は，性格特性を測定するための簡易版診断ツールであり，BIG5（ビッグファイブ）と呼ばれる性格特性を，10 項目の質問を通じて評価するものである．BIG5 の特性の説明を以下に示す．

- **外向性（Extraversion）**：活発さや社交性，ポジティブな感情を示す性格特性である．
- **開放性（Openness）**：知的好奇心や新しい経験に対する関心を示す性格特性である．
- **協調性（Agreeableness）**：他者への思いやりや協力性を示す性格特性である．
- **勤勉性（Conscientiousness）**：責任感や計画性，自己管理能力を示す性格特性である．
- **神経症傾向（Neuroticism）**：ストレスや不安に対する感受性を示す性格特性である．

これらの個人特性は，会話においても重要になってくると予想される．例えば，協調性が高い人は相手に合わせてターンを取る可能性が高いと予想できる．こうした個人特性とターン速度の関係を明らかにすることで，利用者と高い信頼性を築く対話システムの構築などに活かせる可能性がある．

2.5 ターンテイキング検出方法

本研究では，ターンテイキングを発話区間からルールベースに基づき検出する．本手法はターンテイキングの注釈を手動で行う場合と比較し発話区間のみで検出が可能なため，多くのデータの分析を行うのに適している．沈黙を伴うものは先行研究に基づき，オーバーラップを伴うものについては先行研究を参考に定義した [22]．図 1 にターンシフトの検出方法を示す．なお，本研究では *pre-offset* および *post-onset* は 0.3 秒とした．



(a) 沈黙　　(b) オーバーラップ

図 1: ターンシフトの検出方法

1. 区間 $[t_1, t_2]$ の特性に基づく分類:
   - **沈黙を伴うターン交替**: $[t_1, t_2]$ の全ての時刻で両者の音声活動がない（沈黙）．
   - **オーバーラップを伴うターン交替**: $[t_1, t_2]$ の全ての時刻でユーザー 1 とユーザー 2 が同時にアクティブである（オーバーラップ）．
2. t_1 直前の条件（*pre-offset*）:
   - ユーザー 1 がアクティブであること．
   - ユーザー 2 が非アクティブであること．
3. t_2 直後の条件（*post-onset*）:
   - ユーザー 1 が非アクティブであること．

• ユーザー 2 がアクティブであること．

4. 上記の条件を満たす区間をターンシフトと定義する．

3 分析結果

3.1 コーパス全体のターン速度

図 2 は，コーパス全体におけるターン速度の分布を示している．表 1 に，全体の統計情報をまとめた．先行研究で日本語のターン速度が 7 ミリ秒と示されているものと比較し，やや遅めの値が観測されたが，ターン速度分布の形状は先行研究とほぼ同じ概形である [19, 23]．

表 1: コーパス全体のターン速度

指標	全体
サンプル数	3148
平均 (秒)	0.349
分散 ($秒^2$)	0.513
オーバーラップ割合 (%)	29.9

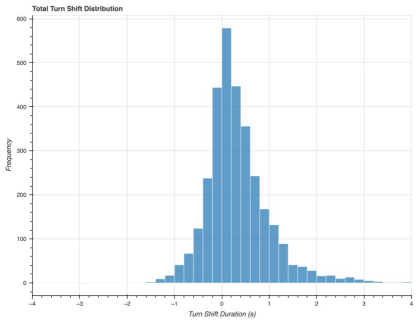


図 2: コーパス全体におけるターン速度の分布

3.2 話者の役割および関係別のターン速度

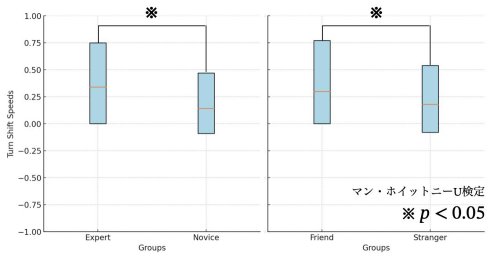
表 2 および図 3 に，話者の役割（初学者から専門家，専門家から初学者）および話者の関係（友人と初対面）に基づくターン速度の統計情報を示す．

初学者から専門家へのターン速度の平均は 0.455 秒，専門家から初学者へのターン速度の平均は 0.247 秒であり，マン・ホイットニー U 検定により有意差（$p < 0.05$）が確認された．また，友人同士の平均ターン速度は 0.462 秒，初対面のターン速度は 0.259 秒であり，こちらもマン・ホイットニー U 検定により有意差（$p < 0.05$）が確認された．

図 3 では，ボックスプロットを用いて各グループ間のターン速度分布を示している．結果として，初学者から専門家，あるいは友人条件においてはより長いターン速度が出現する傾向にあることが明らかになった．

表 2: 話者の役割および関係別のターン速度

指標	初学者から専門家	専門家から初学者	友人	初対面
サンプル数	1543	1605	1391	1757
平均 (秒)	0.455	0.247	0.462	0.259
分散 ($秒^2$)	0.614	0.395	0.637	0.397
オーバーラップ割合 (%)	25.7	34.0	25.7	33.2



(a) 役割別　　(b) 関係別

図 3: 話者の役割および関係別のターン速度

3.3 BIG5 とターン速度の関係

BIG5 の 5 つの性格特性（外向性，開放性，協調性，勤勉性，神経症傾向）に基づくターン速度の統計情報を表 3 に示す．特に，開放性，協調性，勤勉性，神経症傾向において有意差（$p < 0.05$）が観測された．開放性については平均値に大きな差はないものの，高開放性の群では分散が大きい．協調性については，高協調性の群のターン速度が速く，低協調性の群の分散が非常に大きい．勤勉性については，高勤勉性の群はターン速度が速いことがわかった．神経症傾向では，高神経症傾向の群はターン速度が速いことが明らかになった．

図 4 では，各性格特性に基づくターン速度の分布を各群ごとにボックスプロットで示している．外向性では差が小さい一方，他の特性では明確な傾向が観察される．

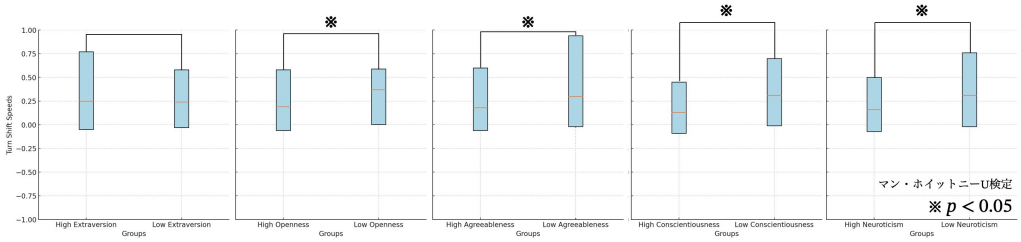
4 考察

本研究の結果から，ターン速度は話者の役割，関係性，個人の性格特性によって大きく異なることが示唆された．これは，対話における発話タイミングが単なる言語的要素だけでなく，心理的・社会的要因に強く影響されることを示している．

話者の役割について: 初学者から専門家へのター

表 3: BIG5 とターン速度の関係

指標	外向性 (高)	外向性 (低)	開放性 (高)	開放性 (低)	協調性 (高)	協調性 (低)
サンプル数	1360	1512	2470	306	1762	424
平均 (秒)	0.430	0.324	0.325	0.362	0.313	0.572
分散 ($秒^2$)	0.681	0.387	0.534	0.285	0.487	0.990
オーバーラップ割合 (%)	29.2	27.9	31.2	25.2	31.6	26.4
指標	**勤勉性 (高)**	**勤勉性 (低)**	**神経症 (高)**	**神経症 (低)**		
サンプル数	727	1636	727	1636		
平均 (秒)	0.233	0.400	0.233	0.400		
分散 ($秒^2$)	0.431	0.465	0.431	0.465		
オーバーラップ割合 (%)	34.8	26.1	34.8	26.1		



(a) 外向性 (b) 開放性 (c) 協調性 (d) 勤勉性 (e) 神経症傾向

図 4: BIG5 の性格特性に基づくターン速度の分布

ン速度が専門家から初学者へのターン速度よりも遅いことは，専門家が議論を主導し，初学者側はより熟考した応答を行う必要性に起因すると考えられる．これは，専門家が提供する情報の正確性や説得力を重視するため，応答の遅延が許容される環境が背景にある可能性が高い．

話者の関係性について: 友人同士の会話では，ターン速度が遅い傾向が見られた．この現象は，友人関係の中で沈黙が許容されやすいという社会的背景に関連していると考えられる．一方，初対面では相手に良い印象を与えたいという心理的プレッシャーがターン速度を速める要因になっている可能性がある．

性格特性の影響について: 性格特性の違いがターン速度に与える影響も興味深い．例えば，協調性のスコアが低く出ている人はターン速度が遅く，分散も非常に大きい．これは，相手の発話に合わせず自己中心的に発話を行う傾向を反映している可能性がある．また，勤勉性に関するスコアが高い人では迅速な応答が見られ，これは責任感や計画性の高さと関連していると考えられる．一方で，神経症傾向に関連したスコアが高い人のターン速度が速いことは，感情の不安定さが対話中の反応速度を高める要因として機能している可能性を示唆している．

5 おわりに

本研究では，ターンテイキングにおけるターン速度の違いを分析し，その要因を明らかにすることを目的として，話者の役割，関係性，性格特性に基づくターン速度の分布を調査した．分析の結果，ターン速度は話者の役割や関係性，個人特性 (BIG5) に関連があることが示唆された．

これらの知見は，ターンテイキングモデルの設計において，ユーザーの特性や状況に応じた柔軟なターン速度の調整が重要であることを示している．本研究では，ターン速度の違いがターンテイキングモデルの性能にどのような影響を与えるかについては検討していない．今後の課題として，これらのターン速度の違いをターンテイキングモデルに反映させ，その効果を検証する必要がある．

本研究の成果は，自然な対話を実現するための基盤となり，対話システムや人間-ロボットインタラクションの応用に貢献することが期待される．

謝辞

本研究の一部は科研費 23K24910 の助成を受けて実施した．本研究は理研の大学院生リサーチ・アソシエイト制度の下での成果である．

参考文献

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

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

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

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

[5] Frederick D Erickson. Conversational organization: Interaction between speakers and hearers, 1984.

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

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

[8] Fred Cummins. Gaze and blinking in dyadic conversation: A study in coordinated behaviour among individuals. **Language and Cognitive Processes**, Vol. 27, No. 10, pp. 1525–1549, 2012.

[9] Kazuyo ONISHI, Hiroki TANAKA, and Satoshi NAKAMURA. Multimodal voice activity projection for turn-taking and effects on speaker adaptation. **IEICE Transactions on Information and Systems**, Vol. advpub, p. 2024HCP0002, 2024.

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

[11] Shinya Fujie, Hayato Katayama, Jin Sakuma, and Tetsunori Kobayashi. 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**, pp. 3771–3775. International Speech Communication Association, 2021.

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

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

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

[15] Kobin H Kendrick, Judith Holler, and Stephen C Levinson. 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**, Vol. 378, No. 1875, p. 20210473, 2023.

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

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

[18] Ryo Ishii, Xutong Ren, Michal Muszynski, and Louis-Philippe Morency. 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**, pp. 131–138, 2021.

[19] Tanya Stivers, Nicholas J Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan Peter De Ruiter, Kyung-Eun Yoon, et al. Universals and cultural variation in turn-taking in conversation. **Proceedings of the National Academy of Sciences**, Vol. 106, No. 26, pp. 10587–10592, 2009.

[20] Atsushi Oshio, ABE Shingo, and Pino Cutrone. Development, reliability, and validity of the japanese version of ten item personality inventory (tipi-j). **Japanese Journal of Personality/Pasonariti Kenkyu**, Vol. 21, No. 1, 2012.

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

[22] Erik Ekstedt and Gabriel Skantze. Voice activity projection: Self-supervised learning of turn-taking events. **arXiv preprint arXiv:2205.09812**, 2022.

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