

Mind the Gap: Temporal Dynamics of Turn Transition in L2 Interviews

Hae-Sung Jeon¹, Simon Wehrle², Aviad Albert², Francesco Cangemi³, Jane Setter⁴, & Martine Grice²

¹*School of Psychology and Humanities, University of Lancashire, UK*

²*IfL – Phonetics, University of Cologne, Germany*

³*Center for Global Education, University of Tokyo, Japan*

⁴*Department of English Language and Applied Linguistics, University of Reading, UK*

This study examines multimodal signals during turn transitions in L2 English face-to-face interaction by L1 Japanese speakers. Turn transitions are fundamental to interaction, with their temporal dynamics shaped by cultural background, conversational context, pragmatic meaning, and language proficiency. Analysing these dynamics provides insights into how interlocutors process, plan, and coordinate responses [1].

We analysed two open and two polar question-response pairs from unscripted interview data ([2]) involving interviewer-interviewee dyads (university professor and student, respectively) across two proficiency groups (11 beginners: 5 female, 6 male, mean age = 18.64 years, with A2 CEFR levels and TOEIC scores between 200-400; 13 intermediate learners: 9 female, 4 male; mean age = 20.68 years, with B2 CEFR levels and TOEIC scores between 800-900). The analysis examined the timing of multimodal signals using the 4V scheme—Visual (visible meaningful movements), Vegetative (non-speech sounds preparing for speech production, e.g., in-breaths), Vocal (non-lexical speech sounds, e.g., filled pauses), and Verbal (lexical responses)—following Conversation Analysis (CA) principles of interlocutors' sequential handling during turn-taking [3]. Three trained annotators manually annotated the video and audio data using ELAN [4] and Praat [5], with annotations cross-checked for reliability. Bayesian mixed models tested response timing from question offset, with proficiency and question type as fixed effects and speaker and item as random effects.

While CA traditionally treats vocal production as the “response” and preceding visual signals as “pre-beginnings” [3], our findings reveal more complex patterns. First, although L2 speakers showed approximately 200 ms latency for one of the 4V signals—consistent with L1 cross-linguistic norms ([6])—they deployed visual and vegetative signals before verbal responses were produced (Fig. 1). We interpret this as prioritising interactional expectations by signalling a willingness to respond within the typically rapid turn transition time window. Second, contrary to previous claims that low L2 proficiency may delay vocal signal timing (e.g., [7]), we found no clear proficiency effect on vocal timing, though substantial individual variation emerged across both proficiency groups. However, there was a robust effect of proficiency on visual signal timing, as confirmed by Bayesian modelling, with beginners responding notably more slowly than intermediate learners (Fig. 2). Further, there was some evidence suggesting a potential interaction between proficiency level and question type.

These findings suggest that, for polar questions requiring simple binary responses, L2 interlocutors maintain the “minimal gaps and overlaps” principle when comprehension is successful. For open questions, however, the greater cognitive demands of comprehension and response formulation reveal proficiency effects even at the “pre-beginning” stage, highlighting the interplay between linguistic competence and interactional coordination using multimodal cues.

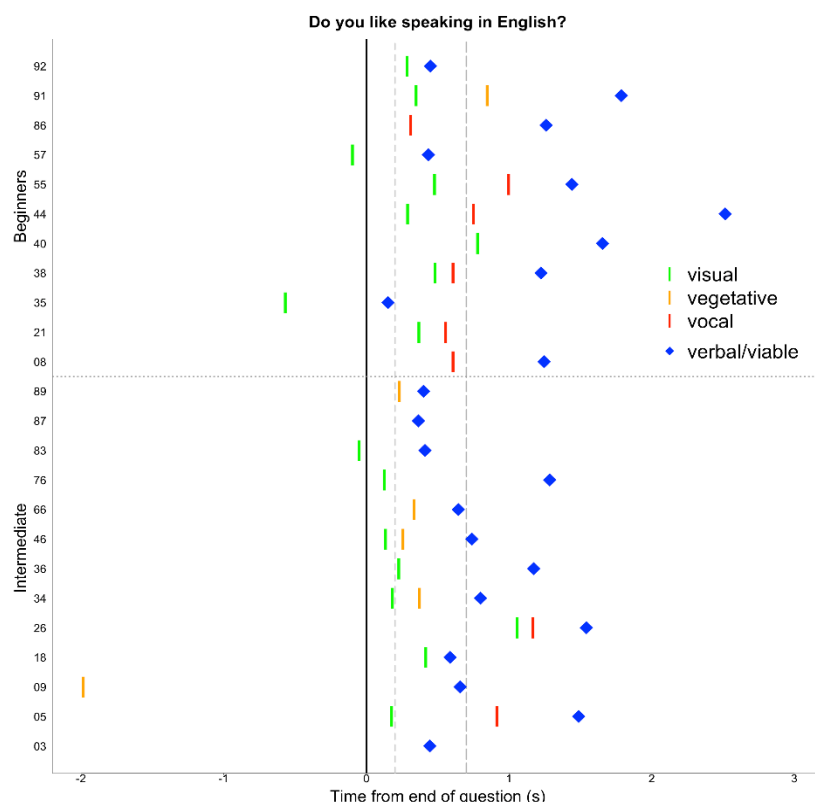


Fig. 1. Time from question offset for 4V signals for a polar question for each participant. Dashed lines mark 200ms (an optimal exchange gap) and 700ms (the upper limit for acceptable gaps) from question offset.

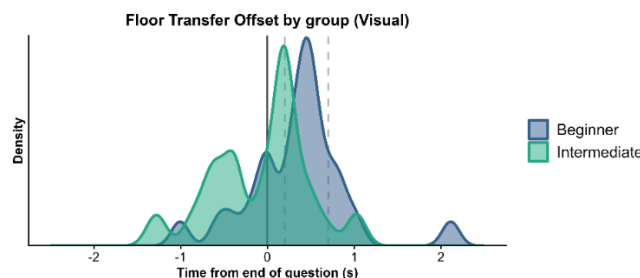


Fig. 2. Density plots for time between question offset and visual signal onset.

References

- [1] Holler, J., Kendrick, K. H., & Levinson, S. C. (2018). Processing language in face-to-face conversation: Questions with gestures get faster responses. *Psychonomic Bulletin & Review*, 25(5), 1900-1908. 10.3758/s13423-017-1363-z
- [2] Ishikawa, S. (2019). The ICNALE Spoken Dialogue: A new dataset for the study of Asian learners' performance in L2 English interviews. *English Teaching (The Korea Association of Teachers of English)*, 74(4), 153-177. 10.15858/engtea.74.4.201912.153
- [3] Schegloff, E. A. (1996). Turn organization: one intersection of grammar and interaction. In E. Ochs, E. A. Schegloff & S. A. Thompson (Eds.), *Interaction and Grammar* (pp. 52-133). Cambridge University Press. 10.1017/CBO9780511620874.002
- [4] Boersma, P. & Weenink, D. (2023). Praat: doing phonetics by computer [Computer program]. Version 6.3.09, retrieved from <https://praat.org>.
- [5] ELAN (Version 6.5) [Computer software]. (2023). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from <https://archive.mpi.nl/tla/elan>
- [6] Stivers, T., Enfield, N. J., Brown, P., & Levinson, S. C. (2009). Universals and cultural variation in turn-taking in conversation. *PNAS*, 106(26), 10587-10592. 10.1073/pnas.0903616106
- [7] Gardner, R. (2007). 'Broken' starts: Bricolage in turn starts in second language talk. In Z. Hua, P. Seedhouse, L. Wei & V. Cook (Eds.), *Language Learning and Teaching as Social Interaction* (pp 58-71). Palgrave Macmillan. doi.org/10.1057/9780230591240_5