The Costs of Fluent Conversation: Speech Planning Reduces Input Processing

Mathias Barthel (mathias.barthel@uni-koeln.de)[1,2]
[1]Department of German Language and Literature, University of Cologne, [2]Leibniz Institute for the German Language, Mannheim

Conversational turn-taking is characterised by a high degree of temporal precision, with interlocutors frequently planning their responses while still comprehending the incoming turn [6, 8, 9]. This overlap between speech comprehension and production has been shown to incur increased processing costs for production [2, 3]. However, it remains unclear whether comprehension processes are likewise impaired during concurrent speech planning, and if so, at which level of linguistic processing [4, 10]. Competing accounts propose either that comprehension is prioritised to ensure accurate input processing, or that speech planning takes precedence to enable timely turn transitions [1, 5].
The present study tests these alternatives by examining whether phonological input processing is reduced during concurrent speech planning. Across four experiments, participants performed phoneme or tone detection tasks while listening to quiz questions that either allowed for response planning during the question (early-planning) or only at the end of the question (late-planning). In Exp. 1a, participants simultaneously monitored the question for a target phoneme and answered the question verbally. Exp. 1b served as a baseline without speech planning. Exp. 2 replaced phoneme detection with tone detection to assess domain-general auditory processing.
Phoneme detection accuracy revealed a significant interaction between planning and task demands (Fig. 1a). In the dual-task condition (Exp. 1a), but not in the control condition (Exp. 1b), phoneme detection performance was significantly reduced in early-planning trials, where response planning overlapped with phoneme perception, compared to late-planning trials, where planning and perception did not overlap, indicating that reduced performance reflects interference from concurrent planning. Crucially, tone detection in Exp. 2 was not significantly affected by planning, suggesting that the interference is specific to phonological processing rather than general auditory attention.
Latency data further support this interpretation (Fig. 1b). In Exp. 1a, but not in Exp. 1b, phoneme detection responses were slower in early-planning trials than in late-planning trials, indicating increased processing difficulty during concurrent planning, demonstrating that concurrent speech planning disrupts phonological input processing rather than merely adding a uniform processing delay.
Additionally, verbal response latencies confirmed that participants engaged in early planning: Responses were faster in early-planning than in late-planning conditions across experiments, indicating that planning was initiated during comprehension whenever possible. These findings show that participants trade off comprehension accuracy for production efficiency in time-critical turn-taking.
Overall, the results provide converging evidence for the *planning-prioritised hypothesis*. When comprehension and production compete for limited cognitive resources, phonological input processing is reduced to facilitate timely response planning. This supports accounts positing shared representational and processing resources between comprehension and production, particularly at the phonological level [e.g., 7]. The absence of comparable effects in tone detection suggests that this interference is domain-specific rather than purely attentional. These findings suggest that the efficiency of rapid turn-taking is achieved at the cost of reduced accuracy in low-level input processing.

The presented evidence indicates that not only is planning less efficient in overlap with comprehension but comprehension efficiency is also reduced during concurrent planning.

References

[1] Barthel, M. (2021). Speech planning interferes with language comprehension: Evidence from semantic illusions in question-response sequences. In *Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue*, pages 1–14, Potsdam, Germany.

[2] Barthel, M. and Rühlemann, C. (2025). Pupil size indicates planning effort at turn transitions in natural conversation. In Zima, E. and Stukenbrock, A., editors, *New Perspectives on Gaze in Social Interaction: Mobile Eye Tracking Studies*, volume 351, pages 188–205. John Benjamins Publishing Company, Amsterdam. https://doi.org/10.1075/pbns.351.07bar.

[3] Barthel, M. and Sauppe, S. (2019). Speech planning at turn transitions in dialog is associated with increased processing load. *Cognitive Science*, 43(7):e12768.

[4] Fargier, R. and Laganaro, M. (2019). Interference in speaking while hearing and vice versa. *Scientific Reports*, 9(1):5375.

[5] He, J., Meyer, A. S., and Brehm, L. (2021). Concurrent listening affects speech planning and fluency: the roles of representational similarity and capacity limitation. *Language, Cognition and Neuroscience*, pages 1–23.

[6] Levinson, S. C. and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology*, 6(731):10–26.

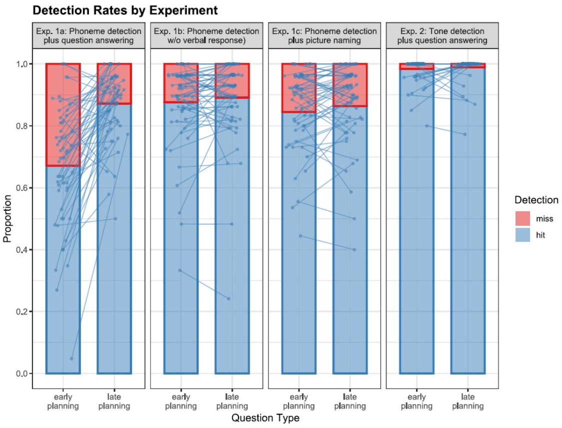
[7] Menenti, L., Gierhan, S. M. E., Segaert, K., and Hagoort, P. (2011). Shared Language: Overlap and Segregation of the Neuronal Infrastructure for Speaking and Listening Revealed by Functional MRI. *Psychological Science*, 22(9):1173–1182.

[8] Pickering, M. J. and Garrod, S. (2013). An integrated theory of language production and comprehension. *Behavioral and Brain Sciences*, 36(04):329–347.

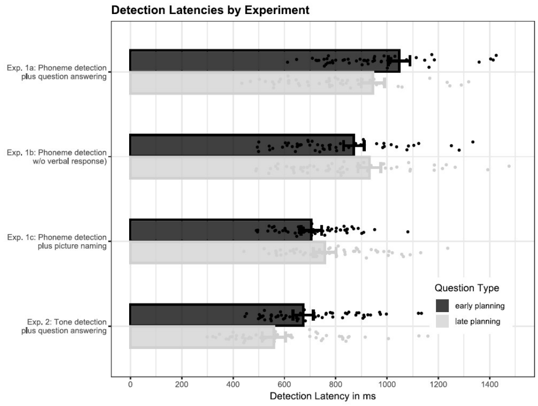
[9] Pickering, M. J. and Strijkers, K. (2024). Language Production and Prediction in a Parallel Activation Model. *Topics in Cognitive Science*, page tops.12775.

[10] Roelofs, A., Özdemir, R., and Levelt, W. J. M. (2007). Influences of spoken word planning on speech recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 33(5):900–913.

Figures and Tables



(a) Detection accuracy by Experiment. Dots indicate mean detection rates by subject. Lines between dots connect subjects' detection rates across conditions.



(b) Model predictions of detection latencies in all Experiments. Error bars represent one standard deviation from the mean. Overlaid dots represent predicted means by participant and condition.