How do listeners time response articulation when answering questions? The role of speech rate

Ruth E. Corps[a], *Chiara Gambi*[b], *& Martin J. Pickering*[a]
[a] Department of Psychology, University of Edinburgh
[b] School of Psychology, Cardiff University

During natural conversation, interlocutors' turns are so finely coordinated that there is often little overlap or gap between their utterances (around 200 ms on average; Stivers et al., 2009). But what mechanisms underlie this striking ability to appropriately time articulation? Speakers often vary in their speaking rates (e.g., Tauroza & Allison, 1990), and so interlocutors must take this information into account if they wish to produce their own turn at the appropriate moment. According to Garrod and Pickering (2015), listeners (as next speakers) determine when they can launch articulation by predicting the speech rate of the speaker's utterance. Specifically, they argued that listeners use the rate of the speaker's preceding syllables to predict the rate of their forthcoming syllables (i.e., they predict the speaker will continue producing syllables at the same rate). This prediction then allows listeners to predict how long it will take the speaker to produce their final syllable, and the moment the listener can take over the turn.

Accordingly, research suggests that speech rate affects syllable comprehension in a way that is consistent with prediction. For example, Dilley and Pitt (2010) found that listeners often failed to perceive a co-articulated single-syllable function word (e.g., *or* in *Deena doesn't have any leisure or time*) when the context surrounding this function word was slowed (*leisure or time* was perceived as *leisure time*). When context rate was speeded, listeners often erroneously perceived an absent function word (e.g., *leisure time* was perceived as *leisure or time*). This effect occurs because the listener uses the speaker's preceding syllable rate to predict that future syllables will be produced at the same rate. This prediction then causes the listener to adopt the interpretation that is more compatible with the predicted rate, leading to the loss or insertion of a syllable (Kösem et al., 2017).

But can interlocutors use these speech rate predictions to coordinate their turns during dialogue? We investigated this issue in two experiments by presenting participants with simple questions (e.g., *Do you have a dog*?) and then instructing them to answer either *yes* or *no*. In Experiment 1, we used time-compression to orthogonally manipulate the context (e.g., *Do you have a…)* and final word (e.g., *dog*?) rate of questions. In particular, we used either a natural rate (normal spoken rate) or a speeded rate (compressed by a factor of 0.5, so it was twice as fast as its natural rate). All questions ended with a monosyllabic final word, which was unpredictable given the context of the speaker's question. To de-confound rate from final word duration, linear mixed effects models with the maximal random effects structure were fitted to answer times from final word onset. Participants (N=32) responded earlier after a speeded (*M*=947ms) than a natural context (*M*=966ms, t = -2.33). Consistent with studies in the speech comprehension literature (e.g., Dilley & Pitt, 2010), this finding suggests that participants used the context rate of the speaker's question to predict that the speaker would produce the rest of their question at the same rate, allowing them to predict when the speaker would reach the end of their question and thus the moment they could launch articulation.

We also found that participants answered earlier after a speeded (*M*=899ms) than a natural final word (*M*=1012ms; *t*=-8.95), regardless of context rate. There are two possible explanations for this effect. First, it is possible that participants adjusted their timing predictions immediately after encountering a final syllable that differed in rate from the question, meaning that they used the rate of the speaker's final syllable to predict when they could time articulation. Alternatively, participants may have responded closer to final word onset when this word was speeded because speeded words are recognized earlier, which in turn allows participants to prepare their own verbal answer earlier. In other words, it is

possible that our final rate manipulation affected answer times because it affected when participants could begin response preparation, rather than because it affects when participants launch answer articulation.

To discriminate between these two hypotheses, Experiment 2 combined our manipulation of final word rate (natural or speeded) with a manipulation of content predictability. In particular, we varied whether response preparation was possible only after the listener had recognized the speaker's final word (unpredictable questions; e.g., *At University, do you study maths?*, like in Experiment 1), or was possible before hearing this word (predictable questions; e.g., *Are dogs your favorite animal*?). By making the final word predictable in half of the questions, we allowed participants to start answer preparation before the rate change occurred and before they even heard the final word. If participants respond closer to the onset of speeded final words because they facilitate earlier response preparation, then we expect an interaction between content predictability and final word rate in Experiment 2. In particular, the final word rate effect should be reduced for predictable compared to unpredictable questions, because response preparation can begin even before participants hear the final word of predictable questions. If, however, the final word rate effect in Experiment 1 was due to rapid adjusting of timing predictions, then we expect the effect of final word rate to be the same, regardless of the predictability of the speaker's final word.

Consistent with previous research demonstrating that listeners use content predictions to prepare a response (e.g., Corps et al., 2018), participants answered earlier when the speaker's final word was predictable (M=665ms) rather than unpredictable (M=947ms; t = -4.63). We also replicated the final word effect from Experiment 1: Participants answered earlier after a speeded (M=748ms) than a natural final word (M=865ms; t = -4.60). Importantly, there was no interaction between content predictability and final word rate (t=0.62, and the Bayes Factor was 0.08, confirming this null interaction), suggesting that our effect of final word rate in Experiment 1 did not occur simply because participants recognized the speaker's final word and began preparation earlier in the speeded than in the natural condition.

Together, these results suggest that listeners can use the speech rate of a speaker's utterance to time initiation of articulation and coordinate their utterances during dialogue. In particular, listeners can form and sustain timing predictions over long timescales (i.e., multiple syllables), but can also adjust these predictions rapidly over shorter timescales (i.e., a single syllable). These findings have important implications for theories of turn-taking, and suggest that timing mechanisms used during language comprehension can influence language production.

References

Corps, R. E., Crossley, A., Gambi, C., & Pickering, M. J. (2018). Early preparation during turn-taking: Listeners use content predictions to determine *what* to say but not *when* to say it. *Cognition, 175,* 77-95.

Dilley, L. C. & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science, 21,* 1664-1670.

Garrod, S. & Pickering, M. J. (2015). The use of content and timing to predict turn transitions. *Frontiers in Psychology, 6,* http://dx.doi.org/10.3389/fpsyg.2015.00751.

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A. S., Jensen, O., & Hagoort, P. (2017). Neural entrainment determines the words we hear. *BioRxiv,* http://dx.doi.org/10.1101/175000.

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., … & Levinson, S. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences, 106,* 10587-10592.

Tauroza, S. & Allison, D. (1990). Speech rates in British English. *Applied Linguistics, 11,* 90-105.