# The Effects of Delay On Task-Based Interactive Conversation

Benjamin Masters,[1*] Ewen MacDonald[1]
[1]Systems Design Engineering, University of Waterloo, Waterloo, ON, Canada
*bpmasters@uwaterloo.ca

Poster 1pSC4

## Introduction

### Turn-taking Dynamics in Conversation

- Measures of turn-taking dynamics such as Interpausal Unit (IPU) durations and Floor-Transfer Offsets (FTO) have been shown to be useful metrics for understanding cognitive processing in conversation [1].
- Behavioral adaptations that result in changes of these metrics, such as a longer and more variable FTO distribution, have been observed when, for example, difficulty is introduced into the conversation by way of noise [2].
- It remains unclear which cues people are using to determine if their speech should be adapted or not.
- This study attempts to examine the sensitivity of talkers to a possible temporal cue, the timing of their partners turn-taking, by manipulating the channel between two talkers to increase response time.
- A previous study has evaluated the effect of delay on conversation in the context of transmission line latency in telephony [3], however the present study extends this analysis by varying delay and by analyzing the results through a modern understanding of conversational dynamics.

### Tracking Conversational State and Manipulating Delay

- It is only necessary to add delay to one channel for it to be perceived by both talkers (see Fig. 1).
- Therefore, the delay is added to the microphone of only one talker, and it is only changed while their partner is speaking to avoid audible processing artifacts that could occur from adding or removing delay during speech.
- The presence of speech was identified on a block-by-block basis using simple voice activity detection based on the background noise level.
- The delay could be manipulated after floor transfers, which occurred when 3 conditions were met
  - The person speaking did not already have the floor
  - The person speaking had been doing so for at least 90 ms, to avoid non-speech acoustic bursts
  - Their partner had not been speaking for at least 180 ms, to avoid misclassifying stop-consonants as pauses [4]
- Delay was manipulated by either padding zeros to the output signal or trimming out silent segments while the other person was talking, as illustrated in Figure 2.

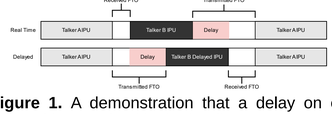


**Figure 1.** A demonstration that a delay on one channel is perceived by both talkers.


**Figure 2.** A signal flow diagram demonstrating the delay manipulation algorithm.

## Data Collection and Processing

### Experimental Design

- 9 pairs of normal hearing native English talkers recruited as friends.
- Performed the DiapixUK tasks in combinations of quiet and noise, and no delay and delay.
- 3 replicates of each condition were completed.
- Delay values were drawn from a uniform distribution with bounds of 0 and 750 ms, chosen to match the FTO IQR increase of normal-hearing talkers in noise vs. quiet.
- Noise used was a synthesized 70 dB SPL babble.

### Data Processing

- Speech activity detected using a VAD threshold 30 dB down from the 99$^{th}$ percentile of power.
- Syllables identified with the Syllable Nuclei v3 Praat script.
- Conversational state classification algorithm used to identify start and end times of IPUs and floor-transfers, ran with both real-time and delayed signals with information from the receiver's perspective kept.
- Statistical analysis performed using general and linear mixed effects models.

## Results

### Floor Transfer Offsets

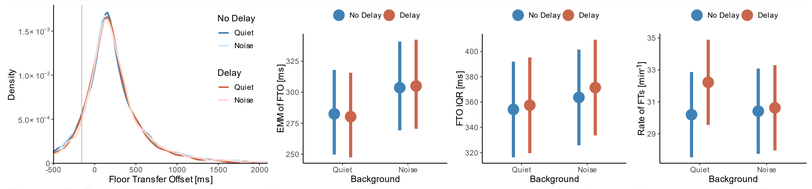


**Figure 3.** Results of FTO analysis. Model results reveal a significant positive effect of Noise on the FTO ($p < .01$), a significant positive effect of Delay on FT rate ( $p < .001$), and a significantly negative effect on the Noise/Delay interaction ($p < .01$). The FT rate is normalized to account for delay.

### Interpausal Units

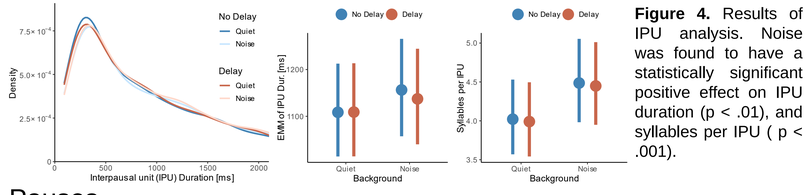


**Figure 4.** Results of IPU analysis. Noise was found to have a statistically significant positive effect on IPU duration ($p < .01$), and syllables per IPU ( $p < .001$).

### Pauses

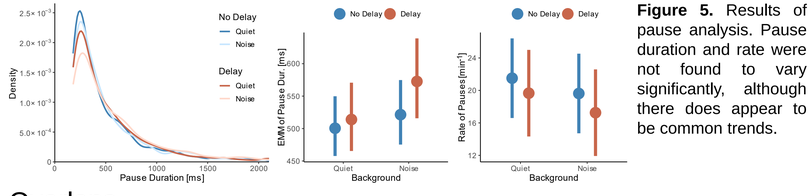


**Figure 5.** Results of pause analysis. Pause duration and rate were not found to vary significantly, although there does appear to be common trends.

### Overlaps

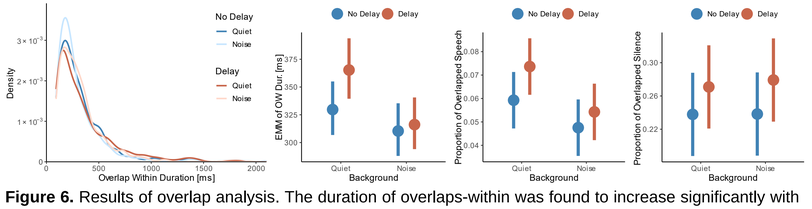


**Figure 6.** Results of overlap analysis. The duration of overlaps-within was found to increase significantly with delay ($p < .01$). The overlapped speaking and non-speaking time, as proportions, both increased significantly with delay ($p < .001$). Overlapped speaking time also decreased significantly with noise ($p < .001$).

## Results, cont.

### Speech Production

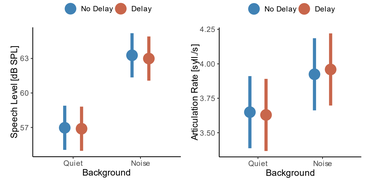


**Figure 7.** Results of analysis of speech features. The mean speech level in each conversation was found to increase significantly in noise ($p < .001$). The conversational mean articulation rate was also found to increase significantly in noise ($p < .001$).

## Discussion

**Noise Results**

- Effects of noise are not as significant as expected, especially in speech level and FTO variability, perhaps suggesting the noise used is not as efficient at masking as expected.
- Despite this, we do see significant effects of noise on the FTO, IPU duration, and speech level and rate.

**Delay Results**

- Delay, when evaluated from the receiver's perspective, had no effect on many of the traditionally evaluated turn-taking metrics.
- However, delay was found to have a significant effect on the periods of overlap in conversation.
  - Overlaps-within became longer.
  - A higher proportion of the conversation was spent with neither person talking.
  - A higher proportion of the conversation was spent with both people talking.
- Delay also significantly affected the rate of floor-transfers.
  - The normalized floor-transfer rate increased, indicating talkers turns shorten.
- Although delay did not affect the FTO, IPU, or Pause durations, it did significantly impact the interactions within the conversation, as evidenced by the increase in overlapped speech and silence.

**Follow up**

- Evaluate masking performance of the noise used.
- Follow up with a different noise or varying noise levels.
- Re-evaluate delay level added as the range was chosen based on the dynamics of NH interacting with HI, which are different than NH-NH interactions.
- Potentially add more delay levels found from FTO variability increases based on hearing status.

## References

1. Levinson, S. C. and Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. Frontiers in Psychology, 6.
2. Sørensen, A. J. M., Fereczkowski, M., and MacDonald, E. N. (2021). Effects of Noise and Second Language on Conversational Dynamics in Task Dialogue. Trends in Hearing, 25:233121652110244.
3. Brady, P. T. (1971). Effects of Transmission Delay on Conversational Behavior on Echo-Free Telephone Circuits. Bell System Technical Journal, 50(1):115–134.
4. Heldner, M. and Edlund, J. (2010). Pauses, gaps and overlaps in conversations. Journal of Phonetics, 38(4):555–568.
5. De Jong, N. H., Pacilly, J., and Heeren, W. (2021). PRAAT scripts to measure speed fluency and breakdown fluency in speech automatically. Assessment in Education: Principles, Policy & Practice, 28(4):456–476.

## Acknowledgements

- This work is funded by the Natural Sciences and Engineering Research Council of Canada: RGPIN-2021-03085