JASA ARTICLE

Variable delay affects conversational turn-taking behavior in the presence of background noise

Benjamin Masters[a] and Ewen N. MacDonald
Department of Systems Design Engineering, University of Waterloo, Waterloo, Ontario, Canada

ABSTRACT:
Previous studies have observed that the duration and variability of floor-transfer offsets (FTO) increase when communication becomes more difficult, such as in the presence of background noise. Additionally, talkers have been shown to adapt their communication behavior when difficulty is increased only for their conversational partner. This study aims to examine whether changes in the timing of FTOs are utilized as a cue by talkers to determine if their partner(s) are experiencing difficulty in communication. A real-time processing system was implemented to randomly vary the delay in a communication line between two talkers so as to alter the duration and variability of FTOs perceived by both talkers. The findings, based on dyadic conversations taking place in both the presence and absence of background noise, with and without delay, reveal that the manipulation of perceived FTO timing does elicit behavioral changes, but only when background noise is also present. This suggests that, when there is an expectation of difficulty, the timing of FTOs may be used as a cue to infer the difficulty level of a conversational partner.
 https://doi.org/10.1121/10.0039572
(Received 21 February 2025; revised 18 September 2025; accepted 26 September 2025; published online 20 October 2025)
[Editor: Susanne Fuchs] Pages: 3107–3119

I. INTRODUCTION

Conversation is a complex interactive activity involving not only speech production and perception, but also the interaction and adaptation of talkers to each other and to their environment.

Talkers adapt to challenging acoustic environments during conversation in a variety of ways. The well-known Lombard effect (Lombard, 1911; Lane and Tranel, 1971) describes a phenomenon in which talkers adjust their vocal effort (e.g., by speaking louder) in the presence of background noise, effectively increasing the signal-to-noise ratio received by any listeners.

However, more subtle adaptations also occur, such as when talkers lean in towards a conversational partner when conversing in noise. When entire conversations take place in noise, interlocutors lean in, providing a signal-to-noise ratio benefit of up to 3 dB when sitting and up to 9 dB when standing (Miles *et al.*, 2023). However, this effect has been observed even in cases where the acoustic benefit is minimal. In Hadley *et al.* (2019), when the background noise level was varied every 15–25 s, participants leaned in, even though the estimated received level increased by only 0.01 dB per 1 dB of added noise. This observation suggests that some adaptations by talkers may occur habitually or be used as social indicators rather than as accommodations that provide acoustic benefits.

Fluid turn-taking in conversation requires interlocutors to perform a near constant monitoring of social, semantic, and behavioral cues that may indicate they should take a turn soon (Gravano and Hirschberg, 2011; Brusco *et al.*, 2020). In an effort to study turn-taking behavior, metrics aimed at quantifying the dynamics of turn-taking have been introduced (e.g., Heldner and Edlund, 2010; Levinson and Torreira, 2015). Notably, floor transfer offsets (FTOs) are defined as the amount of time it takes for one talker to begin their turn after (or before) another talker has ended theirs, interpausal units (IPUs) are defined as connected speech by one talker, and pauses are gaps in a talker's speech (i.e., the silences between IPUs from the same talker where no floor transfer occurs). Figure 1 illustrates a sample dialogue exchange between two talkers, with these metrics labeled. From these features, a conversational turn can be defined as a segment of connected IPUs and pauses by one talker, the starts and ends of which are denoted by FTOs. It is worth noting that FTOs do not occur every time a talker begins speaking, but instead only at instances when there is a floor transfer between talkers. To understand this difference, consider the second IPU produced by talker A in Fig. 1. Since talker B holds the floor (i.e., continues speaking) throughout the entirety of talker A's IPU, no FTO occurs.

Talkers have also been shown to adapt their turn-taking behavior in conversation in response to increased difficulty. Previous studies have found that the durations and variability of FTOs and the durations of IPUs increase in more difficult conditions, such as in the presence of background noise and when conversing in a second language (Sørensen *et al.*, 2021), or for participants with hearing loss (Sørensen *et al.*, 2024; Petersen *et al.*, 2022). These observations have led to the suggestion that increases in duration and variability of FTOs and IPUs can be interpreted as indicators of conversational difficulty (Sørensen *et al.*, 2021). However, such

[a]Email: bpmasters@uwaterloo.ca

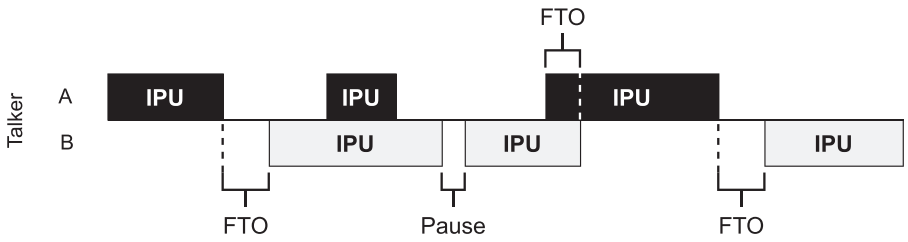

FIG. 1. A sample turn-taking exchange between two talkers illustrating the definitions of and interactions between IPUs, FTOs, and pauses.

interpretations of changes in turn-taking behavior can only be made between conversations that are of a similar nature. It is possible that some types of conversations (e.g., a series of thought-provoking questions) may inherently require more thought or consideration, and therefore contain longer FTOs, than other types (e.g., small talk). Similarly, IPU duration would likely be different between free-form and task-based conversations.

However, talkers do not only adapt in response to the difficulty they experience themselves, but also to the difficulty experienced by their conversational partner(s). Previous work has shown that when one talker in a dyadic conversation received a distorted version of their partner's speech, both talkers exhibited altered speech production, suggesting that the talker receiving the unaltered signal was adapting to the increased difficulty of their partner (Hazan and Baker, 2011). Other studies have examined how speech production and communication behavior change when talkers with normal-hearing (NH) interact with talkers who have a hearing loss (HL). NH talkers have been observed to adapt their speech through increased level, mid-frequency emphasis, and formant frequencies in a manner that was correlated with the level of hearing loss of their HL partners (Beechey *et al.*, 2020). Further, it was found that when NH and HL talkers hold conversations with and without hearing aid amplification, both talkers speak louder when the HL talker is unaided Petersen *et al.* (2022). However, this effect was only observed in conversations taking place in quiet and not in noise, perhaps suggesting that once the NH talker is experiencing increased listening difficulty (as a result of background noise), their sensitivity to the HL partner's difficulty is reduced. It has also recently been observed that NH talkers exhibit significantly different adaptations of turn-taking behavior as an effect of noise when talking with HL partners compared to talking with other NH partners (Sørensen *et al.*, 2024). Although this study does not investigate hearing loss, these findings from previous studies demonstrate that talkers exhibit different adaptations based on the difficulty experienced by their conversational partners.

While the evidence summarized previously suggests that talkers adapt to the difficulty of their interlocutors in conversation, it remains unclear exactly why or how this adaptation is occurring. Given that the previously discussed studies have observed changes in turn-taking behavior as an effect of an expected increase in communication difficulty, one potential explanation is that talkers monitor the timing of their partners' turn-taking to infer their level of effort and adapt accordingly. For example, if talkers perceive that their partner is taking longer to respond than expected and therefore is experiencing difficulty, they may adapt their behavior in an attempt to reduce the amount of effort required by their partner (e.g., by slowing the conversation down to provide their partner with more time to conduct speech understanding and planning). It is also worth noting that there are numerous other potential cues aside from the timing of turn-taking that could be monitored for the same purpose, such as the rate and level of speech produced by a partner, which would need to be investigated separately.

This study aims to examine whether the timing of turn-taking by a conversational partner is used as a cue to infer the difficulty they are experiencing. Specifically, we seek to determine whether talkers alter their speech or conversational behavior when the timing of turn starts by their partner becomes more delayed and more variable (i.e., when their FTOs are longer and more variable). To study this, we recorded interactive conversations between two NH talkers and simulated increases in both the magnitude and variability of FTOs by introducing a variable delay on the communication line between the two talkers. Although it would be more straightforward to implement a constant delay, there is no evidence suggesting that an increase in only the duration of FTOs would be indicative of increased difficulty. As a control to determine how these specific talkers adapt their communication behavior in a difficult environment, conversations were conducted both in the presence and absence of background noise.

It is worth noting that previous studies have investigated how delay impacts communication as a whole, and were motivated with the purpose of determining maximum acceptable transmission line delays in telecommunication systems, such as long-distance telephone lines by Brady (1971), and more recently in digital communication systems, where delay has been studied along with the effects of packet loss (Michael and Möller, 2020). Further, the perceived quality of conversation has been evaluated for differing amounts of transmission line delay (International Telecommunication Union, 2003). The work presented here is significantly different. First, the magnitude of the delay is varied within a conversation with the range of possible delay values based on previously observed differences in conversations observed in quiet vs noise. Second, the analysis investigates different metrics of conversational behavior than have been previously reported by studies investigating the effects of delay on conversation.

https://doi.org/10.1121/10.0039572

There are also some considerations with the experimental setup. Given the presence of an audio delay, talkers had to be situated in separate sound booths such that they could not see each other to avoid a mismatch in synchrony between auditory and visual cues. Thus, the experience of the participants was more in line with a telephone call than a face-to-face conversation. Much of the previous research on behavioral adaptations to which the results of the present study are compared was also based on conversations that were not held face-to-face (Sørensen *et al.*, 2024; Sørensen *et al.*, 2021; Hazan and Baker, 2011).

We hypothesize that there are, broadly, two types of effects that delay could have on a conversation. First, the increased FTO durations of talkers resulting from the presence of delay may be perceived by their partners as an effect of increased difficulty being experienced by that talker. This perceived increase in effort could then result in behavioral adaptations being made in an attempt to reduce the difficulty experienced by the partner. We expect that these adaptations will be similar to those that have been observed in previous studies as effects of noise and hearing status (e.g., increased FTO duration, increased IPU duration, and increased duration and decreased rate of pauses; Sørensen *et al.*, 2024). Some of these adaptations, such as increased FTO duration, may reflect the need for increased processing time by the talker themselves. Others, such as increased IPU and pause duration, may be adaptations that are intended to make the conversation easier for the conversational partner, as longer IPUs and pauses allow more time for speech understanding.

The second type of effect that we anticipate that delay could have on a conversation is a disruptive effect. Due to asymmetric feedback resulting from delay (discussed in Sec. II), it is possible that the natural rhythm of conversation could become disturbed, as talkers may not hear their partners respond when they expect. Additionally, the usage of variable delays will inherently increase the variability of FTOs and may decrease the predictability of turn-taking, thereby making it more difficult for a talker to accurately judge when they should respond. We suspect that the effects of this disruption will resemble the findings of Brady (1971) (e.g., an increase in the proportions of mutual speech and mutual silence). Further, we expect that instances when both talkers begin speaking at the same time will be more frequent, due to a misalignment in the perceived state of a conversation, due to the delay. The consequences of delay on the perception of turn-taking are discussed in the following section.

II. PERCEPTION OF DELAY IN CONVERSATION

How delay affects the perception of turn-taking in conversation by talkers can be counterintuitive. To illustrate this, consider a pair of talkers holding a conversation, which we label talkers "A" and "B." With a delay in the communication line, there will be a mismatch in the perceived timing of turns in the conversation. For example, talker B will not hear talker A begin speaking until an interval equal to the delay has passed since talker A actually began speaking. Due to this mismatch, talker A may think that they are quick to respond, whereas their partner, talker B, is taking longer than expected. However, talker B would have the opposite impression.

An interesting observation that arises from the mismatch of perceived timing is that it is only necessary to delay one talker's speech for the effect to be perceived by both. To understand this phenomenon, we will discuss the effects of delay on FTOs. Consider the dialogue exchange presented in Fig. 2. Following the first turn produced by talker B, talker A will begin their turn after some typical amount of time has passed. Thus, the FTO as perceived by talker A is the amount of time between the end of talker B's turn and when talker A begins their own turn. We denote this amount of time as the "produced FTO," as it is produced by the talker that takes the floor (i.e., starting their turn). However, talker B will not hear the start of talker A's turn until an interval equal to the current delay on the line has passed. Therefore, talker B perceives the FTO as the amount of time between when they end their own turn and when they hear talker A begin speaking, which is equivalent to the sum of the produced FTO and the amount of delay on the line. We denote the FTO as perceived from the perspective of the talker ceding the floor (i.e., ended their turn) as the "received FTO." Thus, talker B perceives the delay at this turn-transition in the form of a lengthened FTO.

Now consider the subsequent FTO, which occurs after talker A ends their turn. Once talker A has ended their turn, some amount of time will pass while talker B is still listening to talker A's delayed speech, after which they will produce some typical FTO. Thus, the delay is perceived by talker A after they have stopped talking, but while talker B is still listening to their delayed speech. Therefore, at this turn-transition, talker A experiences a received FTO as they ceded the floor, and talker B produces an FTO upon beginning their following turn. Note that the received FTO will always be equal to the produced FTO plus the amount of delay present on the line, as visualized in Fig. 2. Thus, while one could delay both microphone signals, the effect on the FTO is the same as applying the sum of both delays to only one microphone signal.

When considering FTOs in conditions with delay, we will use the perspective of the produced FTO, unless explicitly stated otherwise. This is because the produced FTO directly reflects the behavior of the talker who took over the floor, whereas the received FTO is the produced FTO plus a randomized amount of delay. Thus, in the delay conditions, the produced FTO distributions will be analyzed for our hypothesized adaptations (i.e., increased duration and variability).

III. METHODS

A. Participants

Sixteen pairs of young undergraduate participants were recruited as friends and screened for normal hearing

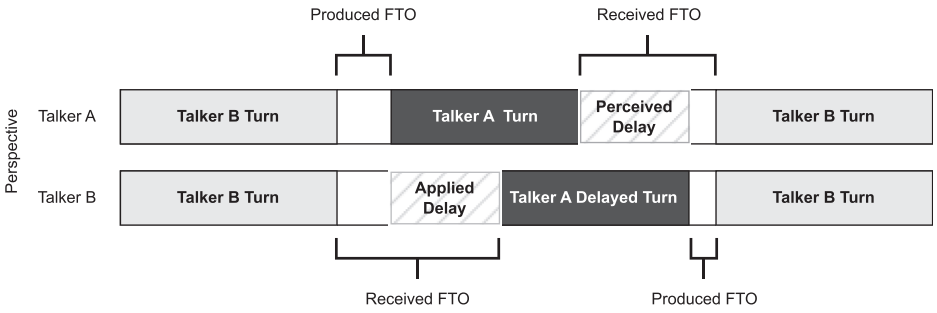

FIG. 2. Demonstration that applying to delay to only one signal leads to a perceived delay by both talkers. Here, the produced FTO is the FTO perceived by the talker that takes over the floor, whereas the received FTO is that perceived by the talker that ceded the floor.

(<20 dB HL) in the frequency range of speech (250–4000 Hz) using an Interacoustics AD226 diagnostic audiometer (Interacoustics, Middlefart, Denmark). Participants self-identified as native English speakers with no history of speech, language, or hearing disorders. The study received ethics clearance from the University of Waterloo Human Research Ethics Board (Ref. No. 46296). All participants provided informed consent, and were remunerated for their time.

B. Setup and equipment

Participants were seated in adjacent sound booths and could not see each other. Each participant wore a headset microphone (DPA 4088; DPA, Kokkedal, Denmark) and a pair of headphones (Sennheiser HD 650, Sennheiser, Wedemark, Germany). One microphone from a matched pair of iSEMCon EMX-7150 measurement microphones (iSEMCon, Viernheim, Germany) was placed in each room, approximately 1.4 meters away from the participants' seating positions, and calibrated with a 94 dB sound pressure level (SPL) 1 kHz test-tone. The headphones' output gains were calibrated to a known dB SPL output level using a GRAS 45CA headphone test fixture (GRAS, Holte, Denmark). The headset microphone gains were set for each participant such that their partner would hear them at the same sound pressure level as if they were seated 1.4 meters away. Talkers always received non-delayed feedback of their own voice through the headphones with unity gain.

C. Experimental conditions

Participants were asked to hold a series of 5 min long free-form conversations. If needed, a list of potential topics was provided to initiate discussion. The conversations took place in conditions that were combinations of quiet vs noise and no delay vs delay. Three conversations were collected per pair in each condition. In the noise condition, the International Collegium of Rehabilitative Audiology normal-effort six-talker babble noise was played through the participants' headphones at 70 dBA SPL (Dreschler *et al.*, 2001). In the delay condition, delay values were randomly sampled from a uniform distribution with bounds of 0 to 750 ms and set to update at every other floor transfer of the conversation. This range of delay values was chosen *a priori* based on empirical observations from a previous study of dyadic conversations taking place between normal-hearing talkers (Sørensen and MacDonald, 2024). To determine an appropriate range, delay values were randomly sampled from a uniform distribution and added to the measured FTOs. The bounds of the uniform distribution were modified until the increase in variance resulting from the delay approximately matched the difference in variance between the conversations taking place in 70 dBA noise versus in quiet. Thus, this range of delays was expected to result in changes to the received FTO distributions that were similar to the differences observed between produced FTO distributions in quiet vs 70 dBA noise when no delay is added.

D. Experimental procedure

In a pre-experimental session, each participant was provided an overview of the experimental setup. Following this, audiometric screening was conducted to ensure that each participant had normal thresholds. This session typically took 15 min to complete.

In the experimental session, each pair of participants was reminded of the experimental setup and told that they may hear background noise in their headphones, but that they should continue to communicate despite this. The participants were not told that a delay would be applied during some conversations. This session consisted of three blocks of four conversations. Each block contained each of the four conditions, and the order of conditions was randomized for each pair of participants. Each conversation was stopped by the operator after 5 min. Within a block, the next conversation was started as soon as the recording system had been set up for the next condition. Between blocks, participants were given a break of approximately 5 min. If participants took their headphones and microphone off during the break, the headset microphone levels were re-calibrated before starting the next block.

E. The delay system

To simulate the observed increase in magnitude and variability of FTOs that have been previously observed in more challenging communication situations, a system was built to introduce a delay in a communication line between talkers. To vary this delay during communication, a near real-time state machine was developed to track floor transfers in conversation as they happened. Given that this experimental setup is testing the effects of delay, it was important that the system had minimal baseline latency. The system was developed using Python, as described in Masters

https://doi.org/10.1121/10.0039572

(2024), and had a baseline round trip latency of ∼6.8 ms while running the algorithm described in the following. Based on previous studies that have investigated sensitivity to delay, this baseline latency should not have affected the communication dynamics between the two talkers (Stone and Moore, 2005; Stone *et al.*, 2008).

The headset microphone signals from each talker were monitored in an ongoing manner using a buffer size of 128 samples at 48 kHz. Voice activity detection (VAD) was performed on each signal for each buffer by assessing whether the root-mean-square (RMS) value of the samples in the buffer was above an energy threshold set based on the background noise in the booth. Time histories of the VAD and speech signals were stored in separate buffers. The VAD history was used to assess recent speaking activity and detect the occurrence of turn-taking. The speech history was used to enable the output of a delayed version of one talker's speech. The identification of floor-transfers and the manipulation of delay will be discussed separately in the sections that follow.

1. Monitoring floor-transfers

Turn-taking was monitored through the tracking of floor transfers, which were determined to occur when the following three conditions were met:

(1) The VAD signal of the talker taking the floor has been labeled as speech for at least 80% of the last 90 ms.
(2) The VAD signal of the talker ceding the floor has been labeled as non-speech for at least 80% of the last 180 ms.
(3) The talker that is taking the floor does not already have the floor.

The first condition, which determines current activity by a talker, is based on a defined minimum IPU duration. The second condition determines the minimum amount of time passed for a turn to be ceded and is approximately twice the duration of silent intervals that are produced during stop consonants in continuous speech. Both intervals are based on the suggestions in Heldner and Edlund (2010). However, the conditions have been slightly relaxed by requiring only 80% of the buffers within these ranges of time to be identified as speech (for condition 1) or non-speech (for condition 2), due to the inability to perform ongoing bridging of short acoustic bursts (e.g., coughs), and short gaps in speech (e.g., stop-consonants) that would typically be done in a *post hoc* approach (Heldner and Edlund, 2010). The third condition simply ensures that the same floor transfer is not counted multiple times. A key point here is that the first condition requires that the talker taking the floor has been speaking for *at least* 90 ms. This ensures that the algorithm can identify floor-transfers with FTOs that are both positive (i.e., gaps between the turns of the two talkers) and negative (i.e., some overlap of the speech from the two talkers' turns). In the case of a negative FTO, the instance at which the system would detect the floor-transfer is later than when the talker who took over the floor started their turn. However, for this portion of the system, we are only interested in determining if a floor-transfer has occurred rather than when it occurred exactly. Thus, this lag is acceptable.

2. Manipulation of delay

The system only added delay to one of the channels, given that this results in a delay perceived by both talkers, as previously discussed in Sec. II and Fig. 2. From here on, to explain the algorithm that manipulated the amount of delay on the line, a nominal talker "A" and "B" will once again be referred to. The delay was only added to the microphone of talker A. As long as talker A was not actively speaking, the amount of delay could be freely manipulated. Thus, when the floor transferred from talker A to talker B, a new random delay value was drawn and could then be implemented using the process described as follows.

The amount of delay was tracked and manipulated using a pointer (denoted from hereon as the "delay pointer") that referenced the delayed position of talker A's speech at any given time, relative to real-time. As delay was only added to talker A's microphone, the audio output to the headphones of talker A was always the most recent buffer of audio from talker B's microphone. However, the output received by talker B at any given time was the buffer of audio just preceding the delayed pointer, accessed via the time history of talker A's speech.

At the start of a conversation, the delay pointer always equaled the real-time position (i.e., there was no delay). However, whenever a new delay value was randomly drawn, the delay pointer was updated.

If the new randomly drawn delay value was greater than the current delay, then delay needed to be added to the line. To add delay, the delay pointer was withheld from advancing (i.e., it was decremented relative to the real-time position) until the difference between the delay pointer and real-time was equal to the new delay target value. While the pointer was being withheld, zeros were substituted for the output of talker A's microphone (i.e., the output was silent).

If the new randomly drawn delay value was less than the current amount of delay on the line, then some amount of delay needed to be removed. To reduce the amount of delay, the system waited until talker A had been silent for an interval equal in duration to the difference between the current and new delay values, upon which the delay pointer was advanced such that the amount of delay on the line matched the target delay value. Since it was required that talker A be silent during the entire interval over which the pointer was skipped forward, it was guaranteed that no speech from talker A would be lost. Note, however, that no such condition was required for adding delay. As a result, it was possible to add a delay every time the floor transferred from talker A to talker B when the new delay value was greater than the current one. However, in the case of removing delay, if talker B's turn was shorter than the amount of delay that needed to be removed, or if talker A produced speech during talker B's turn, then it was possible for there

to be a turn-taking exchange or sequence thereof where the delay could not be manipulated. Thus, there are some floor transfers where delay was not varied.

Pilot testing confirmed that the system produced no audible artifacts when varying delay during floor transfers.

F. Data postprocessing

Voice activity detection was performed on the headset microphone signals. First, the power, in dBFS, of 5 ms windows with 1 ms of overlap was computed. A power threshold was set 25 dB down from the 99th percentile of the power distribution observed for each talker in each conversation. Any window with a power greater than this threshold was classified as containing speech. These results were then further processed in the same manner as Heldner and Edlund (2010). Intervals shorter than 90 ms were assumed to be non-speech events, such as tapping or coughing, and re-labeled as not speech. Silent intervals shorter than 180 ms were assumed to be related to stop-consonants and re-labeled as speech. Speech levels were computed by A-weighting the headset microphone signals and computing the power in the same way. Mean conversational speech levels were estimated using an intermediate voice activity signal where the short silences had not yet been bridged, from which the mean of the A-weighted power signal during speech activity was computed. These results were then converted to dBA SPL based on the gain settings that had been used during the recording.

The speech activity signals were run through a conversational state classification algorithm, which identified IPUs, FTOs, and pauses from the pair of voice activity signals in a conversation. IPUs were identified as intervals of speech longer than 90 ms in duration. Floor transfers were identified to have occurred once one talker had been speaking for at least 90 ms and their partner had not been speaking for at least 180 ms. In instances where this condition was met, the FTO was calculated as the interval from when the talker who ceded the floor stopped to when the talker who took the floor started speaking. Other units were then derived from IPUs and FTOs, such as turns, which were identified as the spans between FTOs, and pauses, which were found as gaps in the speech of a talker who had the floor. Additional features within each IPU and FTO were calculated, such as duration and average level of speech. Further, for each conversation, mutual talking time was computed as the proportion of the conversation when the voice activity signals indicated that both talkers were speaking, and mutual silence time reflected the proportion when voice activity indicated that both talkers were silent. Perceived simultaneous starts were identified when a pair of IPUs from each talker in a conversation began within 100 ms of each other.

As discussed in Sec. II, an FTO, in the presence of delay, is perceived differently by the two talkers. To ensure that the analysis of the FTOs was based on the produced FTOs (i.e., relative to the talker taking the floor), the classification algorithm was run twice. From the perspective of the talker whose microphone was delayed, IPUs, FTOs, and pauses were identified from both real-time headset microphone signals. For the other talker, their real-time signal, along with the delayed signal of the other talker, was used as an input for the identification of the IPUs, FTOs, and pauses.

Conversations in which the delay was varied at less than 20% of the *post hoc* identified eligible floor transfers were removed (6.25% of conversations with delay: three in quiet and three in noise).

G. Statistical methods

Analysis across conditions typically used generalized linear mixed-effects models (GLMM) or linear mixed-effects models (LMM) with the interaction between delay and background noise as a fixed effect, and the talker (or pair) and replicate (i.e., the current repetition of a given condition) as random effects. Models were fit using functions from the lme4 package in R (*glmer* for GLMM, *lmer* for LMM). In some models, other continuous fixed effects were included and will be discussed along with the results. If continuous fixed factors were included, they were scaled prior to fitting. Marginal means were estimated from the fit models using the *emmeans* package. For GLMMs, results are interpreted from the model coefficients, and confidence intervals and p-values were computed with the *lmerTest* package using a Wald t-distribution approximation. For LMMs, a type III analysis of variance was performed using Satterthwaite's method, and results were interpreted from the analysis of variance (ANOVA) output. Unadjusted *p*-values are presented in Sec. IV. The collection of all computed p-values underwent a *post hoc* Benjamini-Hochberg correction to control for multiple testing, and the outcomes of the correction are described in Sec. IV G.

IV. RESULTS

A. Delay implementation verification

First, to verify the effectiveness of the delay implementation algorithm, the distributions of received and produced FTOs in the conversations with delay were compared, and clear differences were observed, as shown in Fig. 3(a). The mean delay in each conversation was also evaluated along with the ratio of the number of instances the delay was manipulated to the count of floor-transfers identified by the *post hoc* conversational state classification algorithm, seen in Figs. 3(b) and 3(c), respectively. Of note is that the delay was varied more consistently in quiet than noise. This difference will be considered in the forthcoming discussion. It can also be noted that there are some conversations (2 in quiet, 0 in noise) where the proportion of FTOs where the delay was varied is greater than 1, this is likely an effect of the constraints around real-time implementation of the same VAD bridging approach as was used in the *post hoc* analysis.

B. Floor-transfer offsets

The FTO distributions for each condition were modeled using the *geom_density* function in R, and are displayed in

https://doi.org/10.1121/10.0039572

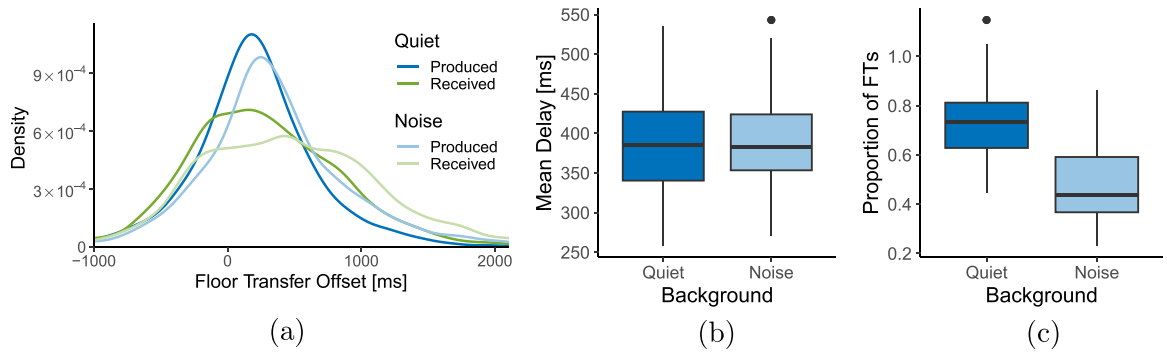

FIG. 3. (a) Distributions of the produced and received FTOs in quiet and noise, in the conversations where delay was present, (b) the mean delay in each conversation, and (c) the proportion of eligible floor transfers at which the delay was varied in each conversation.

Fig. 4(a). It can be observed that the distributions corresponding to the noise conditions are shifted to the right. The delay/noise distribution also appears narrower than the no delay/noise distribution.

The distribution of all FTOs was shifted and truncated such that it only included positive values and was modeled as a gamma distribution. To determine the shift that would result in the best fit, the variance and skewness of all FTOs were computed and used to estimate the mean of the gamma distribution with the same variance and skewness. FTOs that were less than the difference between the estimated gamma mean and the empirical mean of all FTOs were excluded [$\sim$4.54% of all FTOs, <-536.2 ms, indicated by the vertical bar in Fig. 4(a)]. A generalized linear mixed-effects model of the form: FTO $\sim$ Noise $*$ Delay $+$ Pre F.T. IPU Duration $+$ Post F.T. IPU Duration $+$ (1 | Talker Taking Floor) $+$ (1 | Replicate) was fitted to the data using a conditional gamma distribution with a log link function. Here, Pre/Post F.T. IPU Duration corresponds to the durations of the IPUs directly around the floor transfers, which were also included in the model as fixed effects, as the FTO has been shown to depend on these characteristics (Roberts *et al.*, 2015).

The results of this model, which was fitted based on 16 539 FTO observations, revealed a significant increase in FTO duration in noise and with the noise/delay interaction. The model results also showed a significant effect of the duration of the IPU after the floor transfer. Statistical results are summarized in Table I. The estimated marginal means were extracted from the GLMM and corrected for the shift discussed earlier and are displayed in Fig. 4(b).

The variability of the FTOs was measured using the interquartile (IQR) ranges of the FTOs produced by each talker in each conversation ($n = 372$ samples). A linear mixed effects model of the form IQR $\sim$ Noise $*$ Delay $+$ (1 | Talker Taking Floor) $+$ (1 | Replicate) was fitted to the data. An analysis of variance on the LMM revealed a significant positive effect of noise [$F(1, 335.46) = 53.19$, $p < 0.001$, $\eta^2 = 0.14$], but not of delay [$F(1, 335.56) = 0.46$, $p = 0.50$, $\eta^2 = 0.001$], or the interaction [$F(1, 335.46) = 1.59$, $p = 0.21$, $\eta^2 = 0.005$] on the IQR of the FTO. As expected, the FTO distributions become more variable in noise as indicated by a significant increase in the IQR. The conversational FTO IQRs by condition are displayed in Fig. 4(c).

C. IPUs

The IPU duration is a measure of the duration of intervals of continuous connected speech. The IPU duration distributions were estimated by condition using the geom_density function, and are displayed in Fig. 5(a). The plot of the distributions suggests that in noise, the IPU distribution shifts to the right and becomes broader.

The effect of condition on IPU duration was modeled using a generalized linear mixed effects model of the form

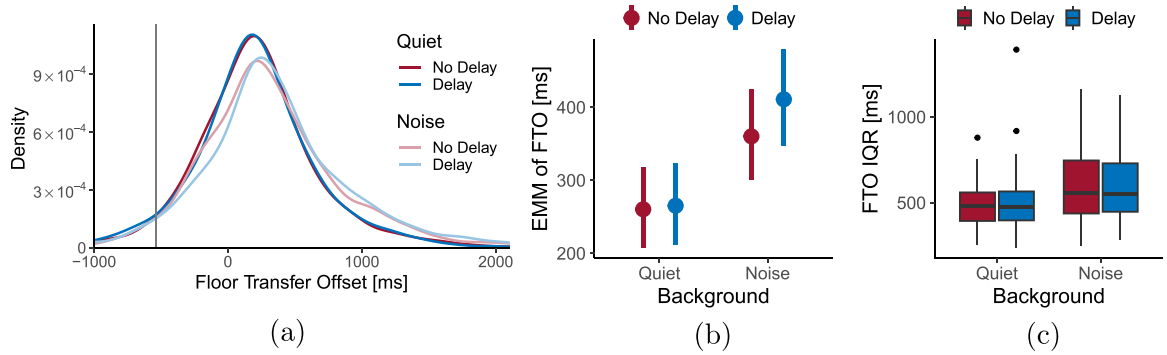

FIG. 4. (a) Distribution of produced floor-transfer offsets by condition, (b) the estimated marginal means of FTO duration, and (c) the interquartile ranges of the FTO. Shading indicates the presence of noise in (a), and the results are grouped by background noise condition in (b) and (c). Color indicates the presence of a delay.

IPU Durations ~Noise * Delay + (1 | Talker) with the same family and link function as was used for the FTO analysis. The inclusion of replicate as a random effect resulted in a singular fit of the model, so it was excluded. Before fitting, the IPUs were shifted by the minimum possible duration of 90 ms such that the shortest measured IPU was 0 ms.

The results of this model, which was fitted based on the observation of 34 220 IPUs, are summarized in Table II. The results revealed a significant positive effect of noise, but not of delay or the interaction of noise and delay, indicating that IPUs become longer when conversing in noise. Figure 5(b) displays the estimated marginal means of IPU duration by condition.

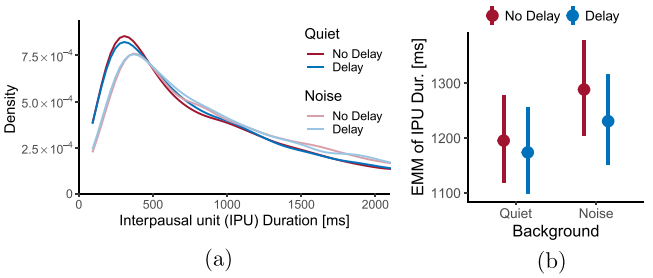

FIG. 5. (a) Distributions of IPU duration by condition and (b) the estimated marginal means of IPU duration. In (a), shading indicates the background noise condition and in (b), the results are grouped by background condition. Color indicates the presence of a delay.

D. Pauses

The duration and rate of pauses can also be indicators of conversational difficulty. The distributions of the durations of pauses by condition were estimated and are shown in Fig. 6(a). The distributions appear to broaden in both noise and with delay.

A generalized linear mixed model of the form: Pause Duration ~ Noise * Delay + Pre Pause IPU Duration + Post Pause IPU Duration + (1 | Talker) + (1 | Replicate) was fit to a gamma distribution with a log link function. Pre/Post Pause IPU Duration corresponds to the durations of the IPUs that immediately surround each pause. Before fitting, the pause durations were shifted by the minimum possible duration of 180 ms, defined by the bridging that occurred during voice activity detection. The durations of the IPUs adjacent to pauses were included as fixed effects in the model to account for the significant increase in IPU duration in noise, as described in 4.3.

The results from the GLMM, which was fitted based on the observation of 16 623 pauses, are summarized in Table III. Significant positive effects of noise and delay were found. A significant negative effect of the pre-pause IPU duration was found, and a borderline positive effect of the post-pause IPU duration was also found. No significant effect of the noise/delay interaction was observed. The estimated marginal means of pause duration are shown in Fig. 6(b).

The rate of pauses in conversation was also analyzed. The rate was calculated by dividing the number of pauses by each talker in each conversation by the sum of the IPU and pause durations of that talker in that conversation ($n = 372$ samples), therefore, no correction for delay was necessary. An LMM of the form: Pause Rate ~ Noise * Delay + (1 | Talker) was fitted to the rates of pauses by each talker. Replicate was excluded as a random effect as it resulted in a singular fit of the model. An analysis of variance on the LMM revealed a significant negative effect of noise [$F(1, 337.16) = 5.97$, $p < 0.05$, $\eta^2 = 0.02$], but not of delay [$F(1, 337.26) = 0.48$, $p = 0.49$, $\eta^2 = 0.001$] or the interaction [$F(1, 337.16) = 0.62$, $p = 0.43$, $\eta^2 = 0.002$]. The conversational pause rates are plotted in Fig. 6(c).

TABLE I. Statistical results for the GLMM fit to the FTO distribution.[a]

Fixed Effect	β-value	Std. Err.	t-value	Pr(>|z|)	
Intercept	6.68	0.04	190.05	< 0.001	***
Noise	0.12	0.01	9.43	< 0.001	***
Delay	0.01	0.01	0.50	0.62	
Pre F.T. IPU Dur.[a]	−0.00	0.01	−0.83	0.60	
Post F.T. IPU Dur.[a]	−0.02	0.00	−5.30	< 0.001	***
Noise:Delay	0.05	0.02	2.69	< 0.01	**

[a]The duration of the IPU directly before/after the FTO.

E. Speaking and listening time

Given the presence of delay, we also expected some changes in the number of instances when talkers began their utterances at the same time, which we denote as perceived simultaneous starts. A perceived simultaneous start was identified when a pair of IPUs (one from each talker) began within 100 ms of each other. In the presence of delay, the number of perceived simultaneous starts may differ between talkers. The total count of these events was computed from the perspective of the talker whose microphone was being delayed (in $n = 186$ conversations). An LMM of the form Perceived Simultaneous Start Count ~ Noise * Delay + (1 | Talker) was fitted to this data, and an analysis of variance was conducted. The results revealed a significant negative effect of noise [$F(1, 167.01) = 22.04$, $p < 0.001$, $\eta^2 = 0.12$] but no significant effect of delay [$F(1, 167.06) = 0.03$, $p = 0.85$, $\eta^2 = 0.0002$] or the interaction [$F(1, 167.01) = 0.04$, $p = 0.84$, $\eta^2 = 0.0003$]. Although this analysis was based on the perspective of only one talker, it was verified that the same pattern of results occurs if the analysis is performed for the other talker in all conversations. The count of perceived simultaneous starts by conversation is displayed in Fig. 7(a).

TABLE II. Statistical results for the GLMM fit to the IPU duration distributions.

Fixed Effect	β-value	Std. Err.	t-value	Pr(>|z|)	
Intercept	7.01	0.04	190.42	< 0.001	***
Noise	0.08	0.01	5.812	< 0.001	***
Delay	−0.02	0.01	−1.43	0.15	
Noise:Delay	−0.03	0.02	−1.48	0.14	

https://doi.org/10.1121/10.0039572

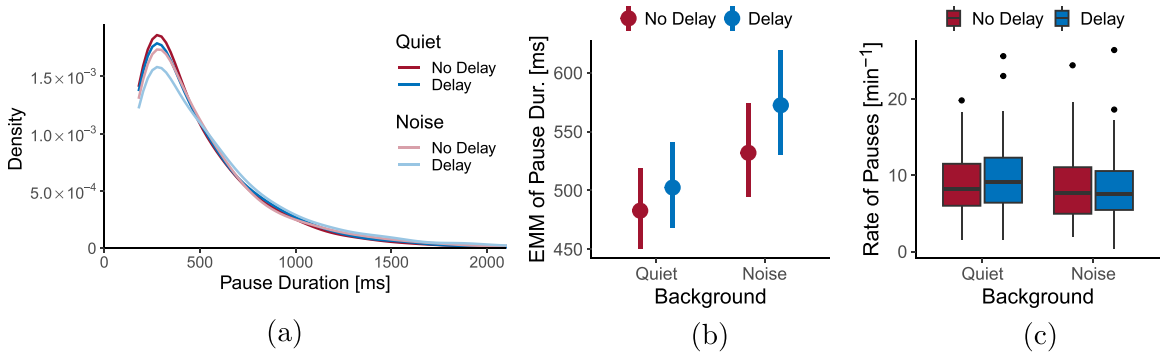

FIG. 6. (a) The distributions of pause duration by condition, (b) the estimated marginal means of pause duration by condition, and (c) the conversational rate of pauses (pauses per minute of speaking time) by condition. In (a), shading indicates background noise condition, and in (b) and (c), the results are grouped by background condition. Color indicates the presence of a delay.

To assess the impact on communication generally, a high level investigation into speech interference between conversational partners was performed. We define mutual talking time as the proportion of the conversation spent with both interlocutors speaking, and mutual silence time as the proportion of the conversation spent with neither speaking. In both cases, to account for the difference in timing of received speech between talkers, the proportions were computed from both talkers' perspectives and then averaged.

An LMM of the form (mutual talking or mutual silence time) $\sim$ Noise $*$ Delay $+$ (1 | Pair) was fitted to the mutual talking and mutual silence time in all conversations ($n = 186$). Analyses of variance revealed the following effects. Mutual talking: a significant negative effect of noise [$F(1, 167.08) = 74.45$, $p < 0.001$, $\eta^2 = 0.31$], a significant positive effect of delay [$F(1, 167.11) = 13.99$, $p < 0.001$, $\eta^2 = 0.08$], but no significant effect of the interaction [$F(1, 167.08) = 2.88$, $p = 0.09$, $\eta^2 = 0.02$]. Mutual silence: significant positive effects of noise [$F(1, 167.03) = 25.06$, $p < 0.001$, $\eta^2 = 0.13$], delay [$F(1, 167.05) = 25.65$, $p < 0.001$, $\eta^2 = 0.13$], and the interaction [$F(1, 167.03) = 4.03$, $p < 0.05$, $\eta^2 = 0.02$]. The conversational mean mutual talking and mutual silence times are displayed in Figs. 7(b) and 7(c), respectively.

F. Speech levels

Characteristics of the talkers' speech were also analyzed. An LMM was fitted to the mean speech level (in dBA SPL) of each talker in each conversation ($n = 372$). The model was of the form: Level $\sim$ Noise $*$ Delay $+$ (1 | Talker). An analysis of variance revealed a significant positive effect of noise [$F(1, 336.99) = 2474.61$, $p < 0.001$, $\eta^2 = 0.88$], but no effect of delay [$F(1, 337.04) = 0.22$, $p = 0.64$, $\eta^2 = 0.0007$] or the interaction [$F(1, 336.99) = 0.03$, $p = 0.86$, $\eta^2 = 0.0001$]. The conversational mean speech levels, in dBA SPL, are plotted in Fig. 8.

TABLE III. Statistical results for the GLMM fit to the durations of pauses.[a]

Fixed Effect	β-value	Std. Err.	t-value	Pr(> |z|)	
Intercept	5.72	0.06	100.14	< 0.001	***
Noise	0.15	0.02	6.43	< 0.001	***
Delay	0.06	0.02	2.72	< 0.01	**
Pre IPU Dur.[a]	−0.02	0.01	−2.86	< 0.01	**
Post IPU Dur.[a]	0.02	0.01	1.94	0.053	·
Noise:Delay	0.05	0.03	1.36	0.17	

[a]The duration of the IPU directly before/after the pause.

G. Statistical correction

The collection of all presented p-values underwent a false discovery rate correction using the Benjamini-Hochberg procedure. With the exception of the interaction effect of noise and delay on mutual speaking time ($p_{adj} = 0.081$), all effects that were significant individually remained significant.

V. DISCUSSION

In this study, we sought to investigate whether talkers monitor the timing of turn-taking to infer the amount of difficulty that their conversational partner is experiencing. By introducing a variable delay on a communication line between two talkers, we aimed to simulate increases in both the magnitude and variability of FTOs that have been observed in previous studies of conversations in conditions with increased difficulty (Sørensen *et al.*, 2024; Sørensen *et al.*, 2021; Petersen *et al.*, 2022). Due to the asymmetry in the way that delay is perceived in conversation (as discussed in Sec. II), while talkers may perceive the FTOs they produce as being typical in duration, they will perceive the FTOs received from their partner as longer and more variable. Thus, the goal of this study was to determine if these perceived changes in the FTOs would be interpreted by talkers as indicators of a partner's difficulty and result in behavioral adaptations in response.

We hypothesized that adding a delay could affect communication in two different ways. The first is that the perceived increase in the duration and variability of FTOs could be interpreted as a marker of difficulty and, therefore, imply increased effort. In this case, talkers may adapt their own speech or behavior in an attempt to reduce the amount of effort that they perceive their partner is exerting. We will refer to these effects as being related to 'perceived effort'. Delay can also be disruptive to the natural flow of

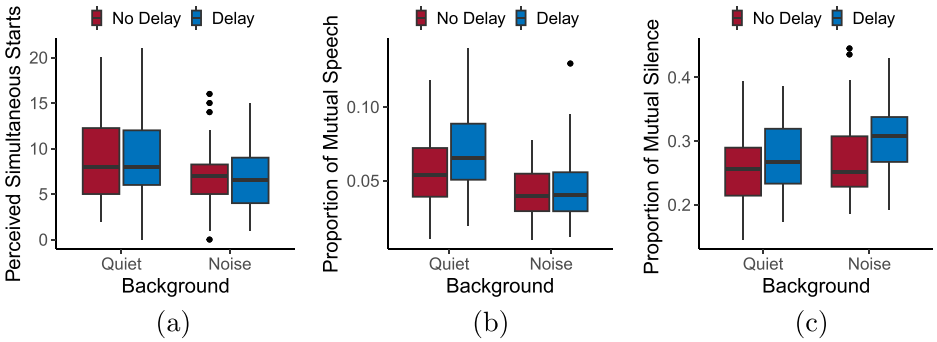

FIG. 7. (a) The count of perceived simultaneous starts in each conversation, (b) the proportion of overlapped speaking time in each conversation, and (c) the proportion of overlapped silent time by both talkers in each conversation. The results are grouped by background noise condition, and color indicates the presence of delay.

conversation. We will consider both of these effects when interpreting our results.

A. Implementation of delay

It is reasonable to consider whether the chosen amount of delay was sufficient to affect communication in the ways we hypothesized. Given that the range of possible delay values was selected based on empirical differences in turn-taking behavior between conversations taking place in quiet versus in the presence of background noise, we expected that similar changes in the FTO distribution would be observed with the addition of delay. Further, a constant delay equal to the midpoint of the selected range of delay values (375 ms) falls near the boundary between the dissatisfaction of "some" and "many" talkers according to telecommunications guidelines (International Telecommunication Union, 2003). Therefore, it seems reasonable to expect communication would be affected in some way by this amount of delay.

Before determining if delay had the anticipated outcomes on turn-taking behavior, it is important to verify whether delay was implemented as expected. As displayed in Fig. 3, it is observed that the mean delay in all conversations is close to the midpoint of the 0–750 ms range of possible delay values. It can also be seen that in all conversations included in the analysis, delay was varied during a substantial portion of the eligible floor-transfers. However, a *post hoc* analysis revealed that delay was varied significantly less often in noise than in quiet [$F(1, 163.16) = 264.77$, $p < 0.001$]. This difference likely arises from the online system's usage of a VAD threshold that was fixed across talkers and conditions. In noise, talkers produce higher speech levels. As a consequence, the proportion of the distribution of speech levels that are higher than the fixed VAD threshold is larger in noise than in quiet. This results in a higher proportion of audio buffers being labeled as containing speech. However, to avoid audible artifacts, the system could only manipulate delays during intervals it detects as being silent. For example, to reduce delay, the system waits for a period of silence equal to the amount of delay that needs to be reduced. The voice activity in the *post hoc* analysis employed an adaptive threshold that was based on the peak levels observed across each conversation and was, therefore, more consistent in determining the boundaries of speech across noisy and quiet conditions. Although the design of the real-time system mitigated this somewhat by slightly relaxing the conditions necessary for identifying turn-taking, it seems from the reduced frequency of delay manipulation that this was not a sufficient accommodation. In future iterations of such real-time manipulation systems, it would be beneficial to use an adaptive VAD threshold to make classification more consistent across different conditions. One option to do this would be to compute such parameters using the power distributions of some speech samples collected at the start of an experiment.

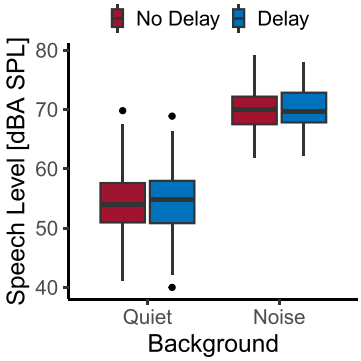

FIG. 8. Mean speech levels (dBA SPL) by conversation. The results are grouped by background noise condition, and color indicates the presence of delay.

Despite this difference in delay variation, the distributions shown in Fig. 3(a) illustrate that the received FTO distributions are broader and right-shifted relative to the equivalent produced FTO distributions. Thus, in both quiet and noise, the implementation of delay had the anticipated outcome on the FTO distributions.

B. Adaptive behavior in response to delay

It has been argued that changes in the FTO distribution may reflect changes in communication difficulty, as an increase in the cognitive demands required to listen will result in less availability of resources to simultaneously plan one's upcoming speech, thereby lengthening the FTO (Sørensen *et al.*, 2021). The results presented here suggest that the reasons for these changes are more complex. The simulated increase in magnitude and variability of FTOs resulted in behavioral adaptations of the timing of turn-taking produced by talkers. Notably, an increase in the produced FTOs was found as an effect of the delay/background

noise interaction. This finding demonstrates that the timing of turn-taking may vary in response to changes in the magnitude and variability of a partner's FTOs, but only in the presence of background noise. We suspect that these adaptations in FTO duration are occurring in an attempt to make the conversation easier, and propose two underlying explanations for their nature. The first is that the adaptations are direct attempts to ease the burden of listening on the conversational partner. By increasing the duration between turns, more time is allocated for the conversational partner to divert their attention and effort away from speech production and toward preparing to listen to and comprehend speech. The second possibility is that the talkers entrained to each other, thus reducing the amount of freedom in the timing of turn-taking behavior and thereby reducing cognitive load (Stel and Vonk, 2010; Levitan and Hirschberg, 2011). Although it is difficult to discern which of these effects is driving the adaptations in FTO duration that were observed, both possibilities indicate that it is, in some way, an adaptation to a perceived increase in difficulty or effort.

A possible explanation for why the FTO duration only increases in the noise/delay interaction and not with delay, in isolation, is that the adaptive behavior may be an effect of an expectation of difficulty that exists in the presence of noise. Thus, a talker may adapt the timing of their turn-taking if they perceive that the increased response time of their partner is an effect of increased difficulty. This suggestion aligns with previous studies that have found that talkers adapt their behavior during communication when increased difficulty is only present for one's conversational partner (Beechey *et al.*, 2020; Hazan and Baker, 2011). Further, given that talkers have been shown to exhibit increases in reciprocal changes in communication behavior as an effect of background noise (Miles *et al.*, 2023), it seems likely that the presence of the effect only in the interaction could be partially attributable to entrainment. We had also hypothesized that, if the delay was perceived as increased difficulty by a partner, talkers would adapt their IPUs by increasing their duration to allow more time for speech understanding. However, no effect of delay or the interaction of noise and delay on IPU duration was observed. We suspect that this may be attributable in part to the delay-induced asynchrony of when speech was produced versus received. This possibility is discussed further in the following section on the potential disruptive effects of delay on conversation.

C. Disruption of communication

Another possible effect of the manipulation of FTOs is the reduced predictability of turn-taking, as talkers may monitor turn-taking dynamics in some serial manner. For example, a talker may expect a shorter IPU to occur after a shorter FTO. Thus, if a manipulation of delay artificially lengthened an FTO that preceded a short IPU, this could disrupt the ability of talkers to accurately monitor the difficulty of their partner over an extended period of time. One possible way to investigate this would be to perform a similar experiment as this study, but instead vary the FTO in some manner that is parametrically related to the durations of the IPU or the turn immediately preceding it. However, there are likely to be challenges in appropriately parameterizing this selection.

Some observations from this study are less likely to be related to adaptive changes as a result of perceived effort. An alternative possibility is that these results are related to communication being somehow interfered with as a result of delay. Additionally, we expected the durations of IPUs to increase if delay was, in fact, perceived as being related to increased difficulty. However, IPU durations did not change as a result of the noise/delay interaction, even though FTO duration did increase. We suspect that this result may also be partially explained by the disruptive nature of the delay. For example, if a talker were attempting to yield the floor but did not hear their partner begin speaking in a reasonable time due to the delay, they may begin speaking again themselves. This would disrupt what would have been the natural flow of the conversation. Other possible disruptions could occur, for example, if a talker begins their speech during a period of silence, but then receives speech from their partner in a delayed manner. In such a case, both talkers would have begun speaking during the same period of silence, but both would perceive the other as interrupting them. One possible effect of scenarios such as these that was observed in this study is the increase in the duration of pauses as a result of delay. Although increased pause duration has been observed in more difficult communication environments, both in this study and in a previous study (Sørensen *et al.*, 2024), we suggest that the effect of delay on pause duration is likely due to disruption, given that no other adaptations were observed as an effect of delay in isolation. If the increased pause duration were an adaptation in response to a perceived increase in a partner's effort or difficulty, one would expect accompanying adaptations such as changes in IPU or FTO durations.

To further analyze the effect that the delay had on communication, some other parameters can be analyzed. It had been previously observed that implementing a constant delay during a conversation, simulating a long distance telephone line, resulted in an increase in both mutual overlapped speaking time and mutual overlapped silence time (Brady, 1971). One suggestion for the interpretation of this finding is that there were more interruptions during the conversations with a delay. This study replicated this finding and extended it by observing that these parameters change as a result of background noise as well, where a decrease in mutual speaking time and an increase in mutual silence were observed. In the context of this study, the replicated findings further support the idea that delay had a disruptive effect on the flow of communication, as it seems more likely that the asymmetric feedback that delay introduced would result in an increase in the frequency of unintentional overlaps than of collaborative overlaps. The additional findings suggest that, in the presence of noise, talkers overlap their

speech less, and a higher proportion of conversations are spent with both interlocutors not talking. This is likely an effect of the FTO distribution shifting towards more positive values, thus lengthening the amount of time between neighboring turns and therefore increasing mutual silence time.

It is also worth pointing out that it is, in general, difficult to determine the nature of such overlaps, IPUs, etc. when voice activity and turn-taking data are analyzed in an automated manner such as that used in this study. A more comprehensive approach to understanding effort, difficulty, and communication breakdowns could incorporate analysis of what is being said, in addition to when it is being said. Such analysis would enable the differentiation of overlaps based on whether they are inherently collaborative, a result of a communication breakdown, or of some other nature. The approach used in this study only evaluated on-off patterns of speech, and the characteristics of speech during portions that were labeled as speech, thus aligning with an existing body of research on conversational dynamics (Sørensen *et al.*, 2024, Sørensen *et al.*, 2021; Petersen *et al.*, 2022; Petersen, 2024). However, it remains difficult to determine with certainty whether some of the results presented here (e.g., increases in mutual speaking or silence time) can be concretely linked to difficulty, disruption of communication, or some other effect.

D. Effects of noise

The background noise condition was included as a control to ensure that some of the conversations taking place were more difficult than others. The changes observed in noise in this study included increased duration and variability of FTOs, increased IPU and turn durations, increased pause duration, and increased speech level. These findings agree with our hypotheses and the results of previous studies, which have compared conversational dynamics between a quiet reference and a high level of background noise condition (Sørensen *et al.*, 2024; Sørensen *et al.*, 2021; Petersen *et al.*, 2022). Given the agreement of results, the background noise appeared to introduce difficulty as expected. It can be observed from the levels of speech in Fig. 8 that talkers tended to increase the level of their speech such that the SNR was, on average, near 0 dB. Although some previous studies have found that talkers communicate in noise with a negative SNR (e.g., –6.3 dB in Petersen, 2024), they involved face-to-face interaction where visual cues that are well-known to be beneficial in challenging conditions were available (Erber, 1975; Sumby and Pollack, 1954). In the present study, if talkers had been face-to-face, the manipulation of acoustic delay would have been obvious. Following the conclusion of the experiment, no participant reported noticing a delay during any conversation.

VI. CONCLUSION

This study evaluated the effect of simulated increases in duration and variability of floor-transfer offsets, implemented by varying the acoustic delay in a communication line. This manipulation was implemented to assess whether talkers use the timing of their partner's turn-taking to infer that difficulty is being experienced. In partial agreement with our expectation, delay was found to increase the duration of FTOs, but only in the presence of noise. This suggests that talkers may use the timing of a partner's turn-taking to infer difficulty and adapt their behavior in response, but this may only be the case when there is an expectation of difficulty. Findings from previous studies on the effects of noise on turn-taking behavior and the effects of transmission line delay on communication were also replicated.

ACKNOWLEDGMENTS

This work was funded by the Natural Sciences and Engineering Research Council of Canada (RGPIN-2021–03085).

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

This study was approved by the University of Waterloo Human Research Ethics Board (Ref. No. 46296). All participants provided informed consent.

DATA AVAILABILITY

The data that support the findings of this study are openly available in Borealis (Masters and MacDonald, 2025). One pair of participants declined the publicization of their data, and as such, are excluded from the public repository, but the data were included in the analysis presented in the manuscript.

Beechey, T., Buchholz, J. M., and Keidser, G. (**2020**). "Hearing impairment increases communication effort during conversations in noise," J. Speech. Lang. Hear. Res. **63**(1), 305–320.

Brady, P. T. (**1971**). "Effects of transmission delay on conversational behavior on echo-free telephone circuits," Bell Syst. Tech. J. **50**(1), 115–134.

Brusco, P., Vidal, J., Beňuš, Š., and Gravano, A. (**2020**). "A cross-linguistic analysis of the temporal dynamics of turn-taking cues using machine learning as a descriptive tool," Speech Commun. **125**, 24–40.

Dreschler, W. A., Verschuure, H., Ludvigsen, C., and Westermann, S. (**2001**). "ICRA noises: Artificial noise signals with speech-like spectral and temporal properties for hearing instrument assessment. International Collegium for Rehabilitative Audiology," Audiology **40**(3), 148–157.

Erber, N. P. (**1975**). "Auditory-visual perception of speech," J. Speech Hear. Disord. **40**(4), 481–492.

Gravano, A., and Hirschberg, J. (**2011**). "Turn-taking cues in task-oriented dialogue," Comput. Speech Language **25**(3), 601–634.

Hadley, L. V., Brimijoin, W. O., and Whitmer, W. M. (**2019**). "Speech, movement, and gaze behaviours during dyadic conversation in noise," Sci. Rep. **9**(1), 10451.

Hazan, V., and Baker, R. (**2011**). "Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions," J. Acoust. Soc. Am. **130**(4), 2139–2152.

Heldner, M., and Edlund, J. (**2010**). "Pauses, gaps and overlaps in conversations," J. Phonetics **38**(4), 555–568.

International Telecommunication Union (**2003**). *ITU-T Recommendation G.114: Transmission Systems and Media: General Recommendations on the Transmission Quality for an Entire International Telephone Connection: One-Way Transmission Time* (Telecommunication Standardization Sector of the International Telecommunication Union, Geneva, Switzerland).

Lane, H., and Tranel, B. (**1971**). "The Lombard sign and the role of hearing in speech," J. Speech Hear. Res. **14**(4), 677–709.

Levinson, S. C., and Torreira, F. (**2015**). "Timing in turn-taking and its implications for processing models of language," Front. Psychol. **6**, 731.

Levitan, R., and Hirschberg, J. (**2011**). "Measuring acoustic-prosodic entrainment with respect to multiple levels and dimensions," in *Proceedings of Interspeech 2011, ISCA*, pp. 3081–3084.

Lombard, E. (**1911**). "Le signe de l'elevation de la voix" ("The sign of raising the voice")," Ann. Maladies l'Oreille Larynx Nez Pharynx **37**, 101–119.

Masters, B. (**2024**). "Talker sensitivity to turn-taking in conversation," Master's thesis, University of Waterloo, Waterloo, Canada.

Masters, B., and MacDonald, E. N. (**2025**). "Replication data for variable delay affects conversational turn-taking behavior in the presence of background noise," Borealis.

Michael, T., and Möller, S. (**2020**). "Effects of delay and Packet-Loss on the conversational quality," in *Fortschritte Der Akustik* (DAGA, Hannover, Germany), pp. 945–948.

Miles, K., Weisser, A., Kallen, R. W., Varlet, M., Richardson, M. J., and Buchholz, J. M. (**2023**). "Behavioral dynamics of conversation, (mis)communication and coordination in noisy environments," Sci. Rep. **13**(1), 20271.

Petersen, E. B. (**2024**). "Investigating conversational dynamics in triads: Effects of noise, hearing impairment, and hearing aids," Front. Psychol. **15**, 1289637.

Petersen, E. B., MacDonald, E. N., and Josefine Munch Sørensen, A. (**2022**). "The effects of hearing-aid amplification and noise on conversational dynamics between normal-hearing and hearing-impaired talkers," Trends Hearing **26**, 233121652211033.

Roberts, S. G., Torreira, F., and Levinson, S. C. (**2015**). "The effects of processing and sequence organization on the timing of turn taking: A corpus study," Front. Psychol. **6**, 509.

Sørensen, A. J. M., Fereczkowski, M., and MacDonald, E. N. (**2021**). "Effects of noise and second language on conversational dynamics in task dialogue," Trends Hearing **25**, 233121652110244.

Sørensen, A. J. M., Lunner, T., and MacDonald, E. N. (**2024**). "Conversational dynamics in task dialogue between interlocutors with and without hearing impairment," Trends Hearing **28**, 23312165241296073.

Sørensen, A. J. M., and MacDonald, E. N. (**2024**). "Metrics conversational dynamics task dialogue by native-Danish talkers across two studies," https://borealisdata.ca/citation?persistentId=doi:10.5683/SP3/FADQIO.

Stel, M., and Vonk, R. (**2010**). "Mimicry in social interaction: Benefits for mimickers, mimickees, and their interaction," Br. J. Psychol. **101**(2), 311–323.

Stone, M. A., and Moore, B. C. J. (**2005**). "Tolerable hearing-aid delays: IV. Effects on subjective disturbance during speech production by hearing-impaired subjects," Ear Hear. **26**(2), 225–235.

Stone, M. A., Moore, B. C. J., Meisenbacher, K., and Derleth, R. P. (**2008**). "Tolerable hearing aid delays. V. Estimation of limits for open canal fittings," Ear Hear. **29**(4), 601–617.

Sumby, W. H., and Pollack, I. (**1954**). "Visual contribution to speech intelligibility in noise," J. Acoust. Soc. Am. **26**(2), 212–215.