INTERSPEECH 2017
August 20–24, 2017, Stockholm, Sweden

Turn-Taking Offsets and Dialogue Context

Peter A Heeman, Rebecca Lunsford

Center for Spoken Language Understanding, OHSU, Portland OR, USA

heemanp@ohsu.edu

Abstract

A number of researchers have studied turn-taking offsets in human-human dialogues. However, that work collapses over a wide number of different turn-taking contexts. In this work, we delve into the turn-taking delays based on different contexts. We show that turn-taking behavior, both who tends to take the turn next, and the turn-taking delays, are dependent on the previous speech act type, the upcoming speech act, and the nature of the dialogue. This strongly suggests that in studying turn-taking, all turn-taking events should not be grouped together. This also suggests that delays are due to cognitive processing of what to say, rather than whether a speaker should take the turn.

1. Introduction

Spoken dialogue systems are starting to allow people to engage in increasingly difficult tasks, in which both the person and system need to contribute to the conversation in an effective way. However, to contribute to a conversation, one needs to have the *turn*. People already know how to engage in turn-taking. Hence, we need to better understand how human turn-taking works so that we can build spoken dialogue system that can engage in turn-taking that is natural for people to use, and is efficient.

In order to build systems that can engage in better turn-taking, there are a number of important questions that need to be answered. At what points in a conversation can turn-taking occur? How is it decided that the current person will stop speaking? How is it decided who speaks next? How quickly is the next speaker expected to start speaking?

Sacks et al. [1] proposed a model that addresses these questions. At certain points, called turn-relevance points, the current speaker can either designate someone to speak or not. If the speaker has designated someone, that person must speak. If no one has not designated someone, anyone can self select; whoever starts speaking first, gets the turn. If neither of the above occurs, the current speaker can continue speaking. Throughout, conversants are trying to minimize gaps and overlaps.

However, there are problems with Sacks' turn-taking model. First, turn-relevance points are not well-defined. Are they defined in terms of pauses, intonation, pragmatics (speech acts), or when the current speaker intends someone else to take the turn? Second, does the other conversant have no control of when a turn-relevance point will occur? Third, is the goal of turn-taking to minimize gaps and overlaps, or are conversants trying to optimize some global measure, like task completion and efficiency, as our earlier work [2] proposed?

In this paper, we build on our recent work in measuring turn-taking offsets [3]. Rather than view turn-taking switches by themselves, we examine them together with turn-taking continuations. Our central question is whether turn-taking offsets are influenced by the dialogue context. We look at local context in terms of what speech act they follow and what speech act they precede. We also examine turn-taking in three different dialogue tasks. We find that turn-taking is influenced by all three factors. This finding calls into question the usefulness of collapsing all types of turn-taking events in studying turn-taking.

This work was funded by the National Science Foundation, IIS-1321146. The authors thank Emma Rennie for helpful conversations.

2. Related Work

Previous work on turn-taking and offsets has aimed to understand what aspects of speech best explain how speakers determine who will speak next, and how speakers manage to, in general, respond so quickly. In addition to the work of Sacks et al. [1], Duncan et. al [4] explored what cues a speaker might use to release the turn, finding that the more cues present, the more likely the speaking turn will change. Gravano and Hirshcberg [5] expanded on this work, also including what cues might invite a back-channel. This work, in essence, takes the viewpoint that the current speaker determines the speaking floor. However, other work has suggested that turn-taking is more collaborative in nature, with each speaker contributing as they can [2, 6, 7, 8].

For analyzing turn offsets, researchers must decide how to segment the speech into turn-construction units. Two common ways of defining them have been silence-based and speech-unit-based. Silence-based segmentation [9, 10, 11] is usually done automatically with a speech detector, and so is not subjective. A minimum within-speaker threshold is used to determine turn-taking continues ranging from 50ms [5] to 200ms [10]. For speech-unit schemes, the speech is segmented based on a single speech act, or a syntactic clause. This segmentation scheme makes it easier to assign pragmatic labels to the units, leading to richer investigations.

A number of researchers have examined factors that might impact turn offsets. Bull and Aylett [12] found that inter-turn offsets change depending on the responders planning needs. Roberts et al. [13] found that turn-taking timing is influenced by both processing needed and speech act sequence, but looked only at question-answer pairs. Stivers et al. [14] also looked at question-answer pairs, finding that turn-taking offsets are quite similar in different languages. Ten Bosch et al. [15] found that turn-exchange offsets differ between phone (shorter) and face-to-face (longer) conversations. These works suggest that offsets are influenced by the preceding speech acts, dialogue context, and cognitive processing needs, although, to date, no research has addressed all three.

3. Data Preparation

The preparation of our data includes segmenting it into turn construction units, annotating the segments, identifying switches and continues, and measuring turn-taking offsets. This is described more fully in our previous work [3].

Corpora: Our data is from three corpora of human-human dialogues. In each corpus, conversants have very different goals, allowing us to determine whether the dialogue task affects turn-

http://dx.doi.org/10.21437/Interspeech.2017-1495

taking behavior. In Trains [16], conversants create a plan for manufacturing and shipping goods. One conversant plays the role of a user, and the other the system, who knows all of the domain constraints, but is told not to drive the conversation. In MTD [17], conversants play a card game where they must collaboratively form a poker hand. Both conversants have identical roles. In Switchboard [18], conversants are told to talk about a certain topic. Although they are not trying to solve a particular task, they do have the goal of carrying on a conversation.

Segmentation: We segment the speech at points where a turn-exchange might have occurred. This includes every point where the speech was semantically, syntactically, and intonationally complete, even if immediately followed by additional speech. Speech was also segmented if incomplete, but provided an opportunity for the other speaker to speak without seeming rude.

Annotation: After segmentation, each segment was annotated as to its speech act (i.e., Inform, Question, Answer, Back-channel, Feedback, Stall, Unknown). Unknown was marked if there was not enough speech to determine its speech act. To help clarify the turn-taking progression, for the segments marked as back-channel or feedback we identified which of the other speaker's segments the speaker was responding to. We also noted segments in which the speaker was talking to himself and segments in which the speaker was clearly intending to interrupt, because in these cases we cannot expect an orderly turn-exchange. Dual-starts were marked where two segments overlap but, unlike an interruption, neither speaker seems to be aware of the other one.

Measurement: Offsets were measured as the length of time from the end of the segment that started immediately before the segment of interest. These are measured for both switches and continues. Interrupts are excluded from our analyses, as we feel that a person who is interrupting is not attempting to follow normal turn-taking behaviors. Self-talk was similarly excluded. Back-channels were measured from the end of the segment that it acknowledged, and both parts of dual-start are measured from the end of the previous segment, with one viewed as a continue and the other as a switch.

4. Turn-taking and Dialogue Task

We first examine whether turn-taking depends on the dialogue task. We examine both the rate at which a segment results in a switch or a continue, as well as the distribution of offsets for switches and continues.

Ratio of Switches vs Continues: We first examine the effect of dialogue task on the percentage of segments that are switches versus continues. The results are shown in Table 1.Using a Chi-squared test of independence, we find a significant effect of task on the number of switches versus continues, X^2 (2, N=6377) = 10.76, p=0.005. When we examine each pair of corpora, we find that the task effect is accounted for by a significant difference between MTD and SW, X^2 (1, N=5084) = 9.44, p=0.002, and a marginal difference between TRAINS and MTD, X^2 (1, N=4772) = 3.60, p=0.058. MTD might have the highest percentage of switches because the task requires a high degree of collaboration, whereas in TRAINS, the system is specifically told not to be too helpful, and in SW, people might share personal anecdotes that take multiple utterances.

Table 1: *Turn-taking ratios and offsets for the 3 Corpora*

		Trains	MTD	SW
Switches	Ratio	49.6%	52.7%	48.0%
	Number	641	1834	771
	Median	0.43	0.29	0.14
Continues	Number	652	1645	834
	Median	0.59	0.22	0.38

Offsets of Switches and Continues: We next examine the effect of corpora on turn-taking offsets. Figure 1 shows the distributions of switches and continues for each of the corpora using box-and-whisker plots, in which the box marks the 1st quartile (Q_1), the median or 2nd quartile (Q_2), and the 3rd quartile (Q_3). Each whisker is 1.5 times the range between 1st to the 3rd quartile (the *interquartile range*), but shortened to correspond to an actual datapoint. The motivation of the whiskers is to denote a 98% region: datapoints beyond this are often considered outliers. We do not show the outliers. The medians for the offsets are shown in Table 1.

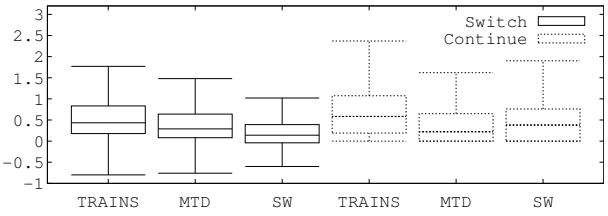

Figure 1: *Offsets and Dialogue Task*

We next determine whether the offsets of switches and continues depends on task. As the median and mean for switches differ for all tasks, as well as for continues, this suggests that the distributions are not normal (skewed to smaller offsets), and so we use the non-parametric Kruskal-Wallis test. We first analyze switches. The median offsets for Trains, MTD, and SW is 0.43s, 0.29s, and 0.14s, respectively, and we find that there is a significant effect of task on offset, df=2, p<.001. Comparing the corpora pairwise, the difference in the offsets is significant for each by Wilcoxon Rank Sum: Trains vs MTD (W=681832, p<.001, two-tailed), Trains vs SW (W=348219, p<.001, two-tailed), MTD vs SW (W=888763, p<.001, two-tailed).

For continues, the median offsets for Trains, MTD, and SW are 0.59s, 0.22s, and 0.38s, respectively, and we find that there is a significant effect of task as well, df=2, p<.001. Comparing the corpora pairwise, the difference in the offsets is significant for each by Wilcoxon Rank Sum: TRAINS vs MTD (W=721796, p<.001, two-tailed), TRAINS vs SW (W=341108, p<.001, two-tailed), MTD vs SW (W=616188, p<.001, two-tailed). Thus we see that dialogue task affects both the rate of switches vs continues, as well as the offset distributions.

5. Turn-Taking and Preceding Speech Act

We next examine the influence of the preceding speech act on turn-taking. It is fairly accepted that the previous speech act affects the rate of switches versus continues. Work on adjacency pairs found that there were common exchanges of speech acts [19], and in fact certain speech acts might create a social obligation for the other conversant to respond, such as a question-answer pair [20]. Some research on predicting turn-taking switches makes use of the previous speech act [21, 22]. While it is accepted that the previous speech act affects the rate of switches, we know of no work that has examined whether the previous speech act affects the distribution of offsets.

Speech Act Rates: As different dialogue tasks might have different rates of speech acts, we start with examining the distribution of speech act types across dialogue tasks (Fig. 2). The rates for each speech act are similar across tasks, but there

are some differences. For example, the relative rate of back-channels in SW is 19.5 times more than in MTD. This might be because grounding in MTD could involve making a next-relevant-utterance rather than a back-channel (cf. [23]). In addition, informs are used 13.7% more absolute in SW than in TRAINS.

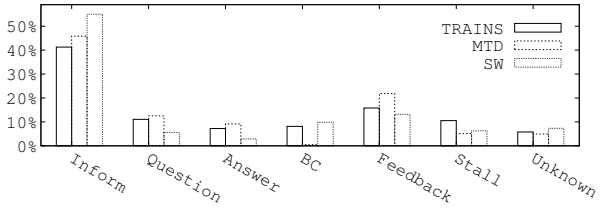

Figure 2: *Speech Act Distribution by Dialogue Tasks*

Ratio of Switches vs Continues: We now examine the percentage at which each speech act type is followed by a switch or a continue. We just use segments that are followed by a turn-taking event. For example, if a back-channel is embedded in an inform, it is not followed by a turn-taking event, and so is excluded.

The results are shown in Fig. 3. As expected, there are differences in the rate of switches versus continues for different speech act types. For example, we see that questions tend to be followed by a switch (81.8% across the three dialogue tasks), presumably because the other speaker answers the question or stalls. We also see that back-channels are typically followed by a switch (94.6%). This is consistent with the use of back-channels to signal understanding and let the original speaker continue. We also see that answers tend to be followed by a switch (77.0%). This is probably because the second speaker is giving feedback after the answer, or because the person who asked the question has dialogue control [24]. Using a Chi-squared test of independence, we found that there is indeed a significant effect of previous speech act on the ratio of switches versus continues, X^2 (6, N=5808) = 830.24, p<.001.

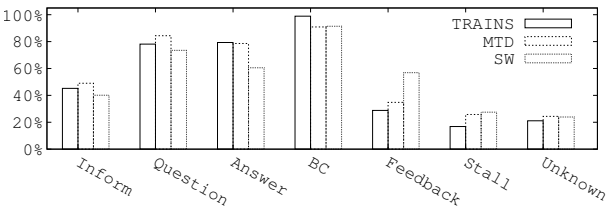

Figure 3: *Ratio that Speech Acts are Followed by a Switch*

In Section 4, we found the ratio of switches versus continues also depends on the dialogue task. We now examine whether the differences between dialogue tasks are just due to the different speech act rates in the corpora, or whether dialogue task has an influence beyond that of the preceding speech act. Using the Cochran-Mantel-Haenszel Test and treating speech acts as a confound, we find that the differences due to dialogue task remain once the effect of the speech acts is accounted for, M^2=10.99, df=2, p=0.004.

Drilling down, we determine which speech acts are affected by the dialogue task in terms of the rate of switches versus continues. Here we use a Bonferroni adjustment to account for multiple comparisons, and the Fisher exact test. We find that there is a significant effect for feedback and informs, p=0.003 and p=0.003, respectively. For the remaining speech acts, we found no effect of dialogue task. Figure 3 shows that, of the speech act types, feedback does have the greatest difference between the three dialogue tasks. For informs, even though the difference between tasks is not as large as other speech acts, there is a lot of data for them, making it easier to reach significance.

Offset of Switches and Continues: We now examine the distribution of offsets for switches and continues following each speech act type. Figure 4 shows the results. As can be seen, different speech acts have different distributions for offsets.

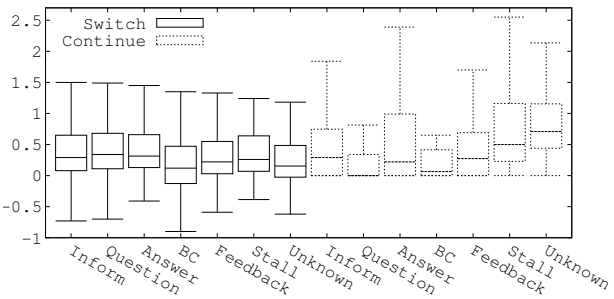

Figure 4: *Offsets by Preceding Speech Act*

Using the Kruskal-Wallis test, we find that there is a significant effect of the previous speech act type on the offsets for switches, df=6, p<.001. Drilling down, we examine the speech acts pairwise to determine whether their switch offsets are different. We used the Wilcoxon Rank Sum with a Bonferroni adjustment. The adjusted p values for all pairs are given in Table 2. Back-channels differ from all other speech acts except unknowns; feedbacks are different from questions, answers and back-channels; and unknowns are different from answers.

Table 2: *Difference between Switch Offsets*

	Quest.	Ans.	BC	Feed.	Stall	Unk.
Inform	NS	NS	<0.001	NS	NS	NS
Question		NS	<0.001	<0.001	NS	NS
Answer			<0.001	<0.001	NS	0.002
BC				0.001	0.001	NS
Feedback					NS	NS
Stall						NS

We ran the above tests for continues. We find that there is a significant effect of the previous speech act type on the offsets for continues, df=6, p<.001. Table 3 shows the results for each pair of speech acts. We exclude back-channels since there are only 12 of them. Here we see that the offsets for continues for most speech act types differ from each other.

Table 3: *Difference between Continue Offsets*

	Quest.	Ans.	Feed.	Stall	Unk.
Inform	<0.001	NS	NS	<0.001	<0.001
Question		<0.001	<0.001	<0.001	<0.001
Answer			NS	<0.001	<0.001
Feedback				<0.001	<0.001
Stall					NS

Interaction with Dialogue Task: Earlier, we found that dialogue task affects the distribution of offsets of switches and continues. It might be that dialogue task has an affect as each task has a different distribution of speech act types. Hence, we now control for the previous speech act type and determine whether dialogue task still has an effect. We use the Kruskal-Wallis test to determine whether there is a significant effect of the dialogue

Table 4: *Effect of Dialogue Task on Offsets*

	Switch		Continue	
	Num.	P value	Num.	P value
Inform	1342	<0.001	1594	<0.001
Question	547	NS	122	NS
Answer	345	0.001	103	NS
BC	209	<0.001	12	too few
Feedback	345	<0.001	575	0.004
Stall	94	NS	310	<0.001
Unknown	49	NS	161	<0.001

task on the offsets for each speech act and a Bonferroni adjustment. The results are shown in Table 4. We find that dialogue task does have an effect for most of the speech act types.

6. Turn-Taking and Next Speech Act

We next examine the influence of the next speech act on turn-taking. Some speech acts tend to follow a certain speech act by the other speaker, as evidenced by adjacency pairs. For example, questions tend to be followed by answers; so switch offsets following a question will be similar to switch offsets before an answer. But the previous speech act does not completely determine what speech act will follow for switches or for continues. For example, after a question, the other person might first make a stall, or ask a clarifying question.

Ratio of Switches versus Continues: We now examine the ratio that each speech act is preceded by a switch versus a continue. Figure 5 shows the results for the three dialogue tasks. As expected, there are differences in the rate of switches for different speech act types. For example, we see that most answers follow a switch, same as for back-channels and feedbacks. Using a Chi-squared test of independence, we find that there is a significant effect of next speech act on the rate of switches versus continues, X^2 (6, N=6366) = 1619.18, p<.001.

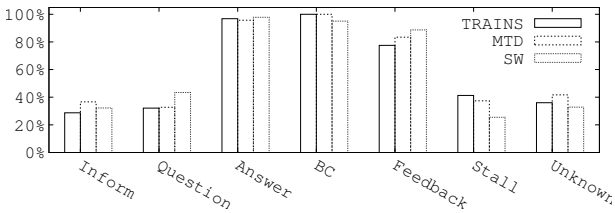

Figure 5: *Rate at which Speech Acts Follow a Switch*

We now examine whether the above differences are just due to the different speech act rates in the corpora. Using the Cochran-Mantel-Haenszel Test and treating next speech act as a confound, we find that the differences due to dialogue task remains even when the effect of the next speech act is accounted for, M^2=8.68, df=2, p=0.013.

Drilling down, we examine each speech act type to determine whether task has an effect on the rate of switches versus continues. Again we use a Bonferroni adjustment to account for multiple comparisons, and the Fisher exact test. We find that there is an significant effect on Informs, p=0.017. For the remaining speech acts, we find no task effect.

Offset of Switches and Continues: Figure 6 shows the distribution of offsets preceding each speech act type for switches and continues. As can be seen, different speech acts have very different distributions for the offsets.

Using the Kruskal-Wallis test, we find that there is a significant effect of the next speech act on the offsets for switches,

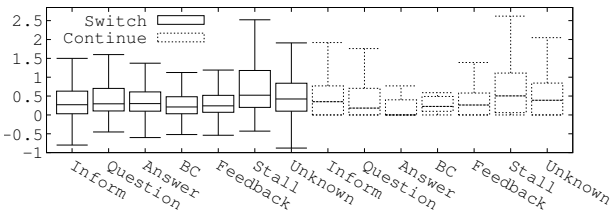

Figure 6: *Offsets by Next Speech Act and Turn-Taking Event*

df=6, p<.001. Drilling down, we examined each pair of speech acts using the Wilcoxon Rank Sum with a Bonferroni adjustment. The results are in Table 5. Stalls differ from everything else, back-channels differ from everything except informs and feedbacks, and unknowns also differ from feedbacks.

Table 5: *Difference between Switch Offsets*

	Quest.	Ans.	BC	Feed.	Stall	Unk.
Inform	NS	NS	NS	NS	<0.001	NS
Question		NS	0.002	NS	<0.001	NS
Answer			0.002	NS	<0.001	NS
BC				NS	<0.001	<0.001
Feedback					<0.001	0.002
Stall						NS

We also ran the above tests for continues. We find that there is a significant effect of the previous speech act type on the offsets for continues, df=6, p<.001. We next examined each pair of speech acts. Stalls differed from informs, questions, and back-channels, but no other differences were significant.

Interaction with Dialogue Task: Earlier, we found that dialogue task affects the distribution of offsets of switches and continues. Here, we control for the next speech act type and determine whether task still has an effect. We again use the Kruskal-Wallis test on each speech act type and use a Bonferroni adjustment. The results are shown in Table 6. We find dialogue task does have an effect for some of the speech acts.

Table 6: *Effect of Dialogue Task on Offsets*

	Switch		Continue	
	Num.	P value	Num.	P value
Inform	1004	<0.001	1959	<0.001
Question	228	NS	444	<0.001
Answer	471	NS	19	too few
BC	279	<0.001	8	too few
Feedback	970	<0.001	194	0.029
Stall	154	NS	276	<0.001
Unknown	138	NS	228	0.001

7. Conclusion

In this paper, we showed that whether turn-taking occurs, and the turn-taking offsets, depends on the dialogue task, and the preceding and upcoming speech acts. We also showed that when accounting for the preceding speech act or upcoming speech act, that the effect of dialogue task is still significant. This work shows the limitation of combining all switch offsets together, as well as all continue offsets together. This work also suggests that in turn-taking, conversants might not be trying to minimize gaps and offsets, but are responding when and if they have something to contribute to the conversation.

8. References

[1] Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson, "A simplest systematics for the organization of turn-taking for conversation," *Language*, vol. 50, no. 4, pp. 696–735, Dec. 1974.

[2] E. O. Selfridge and Peter A. Heeman, "Importance-driven turn-bidding for spoken dialogue systems," in *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, Uppsala Sweden, July 2010, pp. 177–185.

[3] Rebecca Lunsford, Peter A. Heeman, and Emma Rennie, "Measuring turn-taking offsets in human-human dialogues," in *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, San Francisco, Sept. 2016, pp. 2895–2899.

[4] Starkey Duncan, "Some signals and rules for taking speaking turns in conversations," *Journal of Personality and Social Psychology*, vol. 23, pp. 283–292, 1972.

[5] Agustín Gravano and Julia Hirschberg, "Turn-taking cues in task-oriented dialogue," *Computer Speech & Language*, vol. 25, no. 3, pp. 601–634, July 2011.

[6] Stephen J. Cowley, "Of Timing, Turn-Taking, and Conversations," *Journal of Psycholinguistic Research*, vol. 27, no. 5, pp. 541–571, Sept. 1998.

[7] Daniel C. O'Connell, Sabine Kowal, and Erika Kaltenbacher, "Turn-taking: A critical analysis of the research tradition," *Journal of Psycholinguistic Research*, vol. 19, no. 6, pp. 345–373, Nov. 1990.

[8] R. J. D. Power and M. F. Dal Martello, "Some criticisms of Sacks, Schegloff, and Jefferson on turn taking," *Semiotica*, vol. 58, no. 1-2, pp. 29–40, Jan. 1986.

[9] Mattias Heldner and Jens Edlund, "Pauses, gaps and overlaps in conversations," *Journal of Phonetics*, vol. 38, no. 4, pp. 555–568, 2010.

[10] Mattias Heldner, Jens Edlund, Anna Hjalmarsson, and Kornel Laskowski, "Very short utterances and timing in turn-taking," in *Interspeech*, 2011, pp. 2848–2851.

[11] John Kane, Irena Yanushevskaya, Céline de Looze, Brian Vaughan, and Ailbhe N. Chasaide, "Analysing the prosodic characteristics of speech-chunks preceding silences in task-based interactions," in *Interspeech*, 2014.

[12] Matthew Bull and Matthew Aylett, "An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue," in *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP-98)*, Sydney Australia, 1998.

[13] Seán G. Roberts, Francisco Torreira, and Stephen C. Levinson, "The effects of processing and sequence organization on the timing of turn taking: a corpus study.," *Frontiers in psychology*, vol. 6, 2015.

[14] Tanya Stivers, N. J. Enfield, Penelope Brown, Christina Englert, Makoto Hayashi, Trine Heinemann, Gertie Hoymann, Federico Rossano, Jan P. de Ruiter, Kyung-Eun Yoon, and Stephen C. Levinson, "Universals and cultural variation in turn-taking in conversation," *Proceedings of the National Academy of Sciences*, vol. 106, no. 26, pp. 10587–10592, June 2009.

[15] Louis ten Bosch, Nelleke Oostdijk, and Jan de Ruiter, "Durational Aspects of Turn-Taking in Spontaneous Face-to-Face and Telephone Dialogues," in *Text, Speech and Dialogue*, Petr Sojka, Ivan Kopecek, and Karel Pala, Eds., vol. 3206 of *Lecture Notes in Computer Science*, chapter 71, pp. 563–570. Springer Berlin / Heidelberg, Berlin, Heidelberg, 2004.

[16] Peter A. Heeman and James F. Allen, "The Trains spoken dialog corpus," CD-ROM, Linguistics Data Consortium, April 1995.

[17] Fan Yang, Peter A. Heeman, and Andrew L. Kun, "An investigation of interruptions and resumptions in multi-tasking dialogues," *Computational Linguistics*, vol. 37, no. 1, pp. 75–104, Mar. 2011.

[18] J. J. Godfrey, E. C. Holliman, and J. McDaniel, "SWITCHBOARD: Telephone speech corpus for research and development," in *Proceedings of the International Conference on Audio, Speech and Signal Processing (ICASSP)*, 1992, pp. 517–520.

[19] E. A. Schegloff and H. Sacks, "Opening up closings," *Semiotica*, vol. 7, pp. 289–327, 1973.

[20] David R. Traum and James F. Allen, "Discourse obligations in dialogue processing," in *Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics*, Las Cruces, New Mexico, June 1994, pp. 1–8.

[21] Nishitha Guntakandla and Rodney Nielsen, "Modelling turn-taking in human conversations.," in *AAAI Spring Symposium on Turn-Taking and Coordination in Human-Machine Interaction*, Stanford CA, 2015.

[22] Tomer Meshorer and Peter A. Heeman, "Using past speaker behavior to better predict turn transitions," in *Proceedings of the 17th Annual Conference of the International Speech Communication Association*, San Francisco, Sept. 2016, pp. 2900–2904.

[23] Herbert H. Clark and S. E. Brennan, "Grounding in communication," in *Perspectives on Socially Shared Cognition*, L.B. Resnick, J. Levine, and S.D. Behreno, Eds., pp. 127–149. APA, 1990.

[24] Fan Yang, Peter A. Heeman, and Kristy Hollingshead, "Towards understanding mixed-initiative in task-oriented dialogues," in *Proceedings of the 8th International Conference on Spoken Language Processing (ICSLP-04)*, Oct. 2004, pp. 217–220.