Speaker transition patterns in German: A comparison between task-based and casual conversation in face-to-face and remote conversation

Qiang Xia, Marcin Włodarczak
Department of Linguistics
Stockholm University
{qiang.xia,wlodarczak}@ling.su.se

Emer Gilmartin
Inria, Paris
gilmare@tcd.ie

Abstract

The study describes floor transition patterns in free and task-oriented 'spot the difference' conversations by 10 pairs of German native speakers. Each floor transition was delimited by stretches of longer (> 1 s) intervals of solo speech and included an arbitrary number of intervening intervals corresponding to silences, overlaps and shorter stretches of solo speech. While the effect of video conferencing was minor, the type of task had a large effect on the turn-taking patterns. Compared to the free conversation, the task-oriented dialogues were characterised by more frequent speaker changes, particularly short transitions involving a single gap. In addition, within-speaker transitions with three intervening intervals were very common in this condition, especially those in which the interlocutor provided shorter verbal contributions, possibly corresponding to feedback expressions.

1 Introduction

Although widely described as the fundamental mechanism of spoken interaction, turn-taking is still not very clearly understood. Spoken interaction can vary in multiple ways, including number of speakers involved, purpose, register, setting, and medium. It is likely that the temporal arrangement of speech also varies depending on factors such as those mentioned above. In this study, we address this problem by examining the arrangement of contributions by participants in German task-based and free (casual) conversations held face-to-face and remotely over the Internet.

We base our analyses on *floor state dynamics*, where spoken interaction is represented as a series of floor state intervals, describing who is speaking or remains silent at a particular time. The floor state changes constantly throughout the interaction, and sequences of floor states, or *floor state transitions*, capture speech activity patterns, facilitating a data-driven method to analyse the local dynamics of turn-taking in different types of spoken interaction. They can be used to describe turn-taking patterns of arbitrary complexity and provide a convenient starting point for more specific investigations of conversational structure and content.

We perform a within-subject comparison of the floor state dynamics of conversations from a subset of the Berlin Dialogue Corpus (BeDiaCo), version 2 (Belz et al., 2021), where pairs of German speakers engaged in two conversation types (task-free casual conversation, and 'spot the difference' or Diapix tasks) in two sessions – face-to-face and over an internet connection.

2 Background

In this section, we briefly discuss the contextual factors that might condition the emerging patterns of turn-taking in conversation, including the effects of videoconferencing and the organisation of conversation floor, both of which are of direct interest to this study. We also introduce the paradigm used to describe floor transitions used in this work.

2.1 Contextual effects on turn-taking

Even though Sacks et al. (1974) made it abundantly clear that their turn-taking model did not necessarily apply to all speech exchange systems, much of the work on conversational turn-taking adopts the assumption that "overwhelmingly, one party talks at a time" (Sacks et al., 1974, p. 700) as one of the underlying principles of all verbal interaction. However, this is not necessarily the case as the rules governing the temporal arrangement of turns depend on contextual factors such as task, medium and speakers' familiarity (O'Connell et al., 1990).

In particular, Edelsky (1981) demonstrated that in addition to the "one-speaker-at-a-time" model, conversation floor can also be collaborative with several interlocutors engaging in a "free-for-all" state. In a collaborative constructed floor, turn

Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue, September, 11–12, 2024, Trento, Italy.

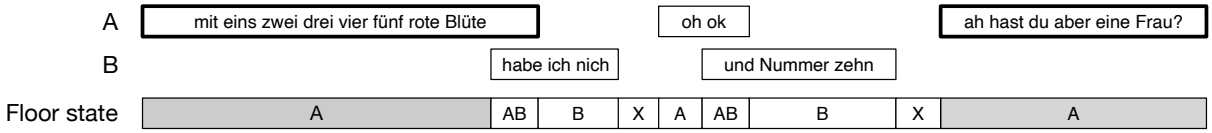

Figure 1: Example of a between-speaker transition. The two top rows represent speakers' talkspurts (*A: With one two three four five red flowers; B: I don't have it; A: oh ok; B: and number ten; A: ah but do you see a woman there?*). The third row represents the floor state with solo-speech intervals longer than one second marked in grey.

length is more evenly distributed compared to a single-floor model, and overlapping speech is considered a sign of participants' active engagement in a shared conversational space. Similarly, Tannen (1980) found that high involvement in conversation is characterized by high speech rate, rapid turn-taking with short gaps and frequent overlaps.

In addition, while our understanding of turn-taking mechanisms and conversational style is predominantly based on face-to-face (and, to some extent, telephone) conversations, the effect of the medium can potentially have a strong effect on temporal patterns of turn exchange. As a case in point, electronically-mediated remote conversations are characterised by an unavoidable electronic transmission delay, which might disrupt the rhythm of conversational turn-taking, causing longer response time in answering polar questions (Boland et al., 2021). Egger-Lampl et al. (2010) found a positive correlation between conversational interactivity and speakers' sensitivity to delay impairments. They demonstrated that in highly interactive telephone conversations, such as random number verification, fewer speaker changes take place under long-delay conditions than under short-delay conditions. This suggests that latency may affect speakers' ability to predict the turn end and they may change their turn-taking behaviours depending on the conversational condition. Indeed, Bailenson (2021) hypothesised that in video conferencing interactions interlocutors need to work harder to send and receive turn-taking cues, which might explain the "Zoom fatigue" reported by some users.

In the present study, we compare speaker transition patterns in conversations characterized by high and low interactivity by contrasting free conversations and the Diapix task (Van Engen et al., 2010; Bullock and Sell, 2022), a spot-the-difference game where participants are each given similar pictures which contain a number of differences and try to find all differences through speech alone. Baker and Hazan (2011) examined Diapix interactions and concluded that it is a valid method for eliciting balanced speech contribution in dyadic conversations. This task allows researchers to analyze conversational dynamics in a controlled but naturalistic setting, providing insights into how participants manage turn-taking in collaborative dialogues. We additionally investigate the effect of the medium by having the same participants conducting both types of interaction face-to-face and using video-conferencing software.

2.2 Analysis paradigm

The analysis of turn-taking patterns in large conversational corpora has a long tradition going back to the seminal work on telephone speech by Norwine and Murphy (1938); Brady (1968); Jaffe and Feldstein (1970), which describes floor transition phenomena in terms of probabilities of solo speech, silence and overlap sequences. This line of research has proven useful for describing temporal properties of turn-taking patterns in interaction (Heldner and Edlund, 2010) and for identifying differences between interactional settings, such as face-to-face and telephone interaction (ten Bosch et al., 2004, 2005). Furthermore, machine learning on speech and silence data from large corpora of dyadic and multiparty speech has been successfully used to infer information about spoken interaction, for example, predicting speaker activity from conversation history (Jaffe et al., 1964; Beebe et al., 1988, 2000), inferring information such as relationships between participants, genre, and features such as personality traits of speakers in dyadic and multiparty interaction (Laskowski, 2011; Gilpin et al., 2018). However, much of this work is built on two assumptions which do not make justice to the complexity of the conversational turn-taking. First, it considers any transition between non-overlapping intervals, however short, as potentially meaningful. Second, it implicitly assumes that speaker change and retention are achieved within a scope of a single interval of silence or overlap.

Consider, for instance, Figure 1, which shows an excerpt from a dyadic conversation. There are nine

Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue, September, 11–12, 2024, Trento, Italy.

floor states – solo speech (three *A*s and two *B*s), overlaps (two *AB*s) and silence (two *X*s). Existing data-driven approaches to turn-taking could treat this stretch as a series of four transitions: two instances of *A_AB_B* from A to B, and two instances of *B_X_A* from B to A. However, looking at the transcript and the speech patterns, it seems more likely that the *longer* stretches of solo speech by speaker A delimit a single more complex transition with A retaining the conversational floor. Such larger conversational structures are routinely overlooked by large-scale corpus studies.

2.3 Floor state transitions

A more detailed approach to describing floor transitions like those in Figure 1 was proposed in Gilmartin (2021). In this approach, longer sequences of speech and silence were captured by concatenating floor state intervals. Floor state transitions were identified as the sequence of intervals between stretches of solo (single-party) speech in the clear (without overlap), in order to gain insight into how turn change and retention is managed by participants. To approximate turn changes or retention, a minimum duration threshold was placed on the solo speech intervals leading into and out of the transitions. Transitions were classified as within- or between-speaker (WST and BST, respectively), depending on whether speaker change occurred or whether the same speaker continued.

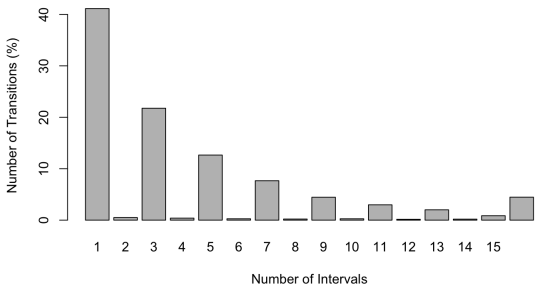

Figure 2: Frequency of transitions with different numbers of intervening intervals in a corpus of casual (free) conversation, reproduced from Gilmartin (2021).

This approach was used by Gilmartin et al. (2020), who identified turn transitions in 24 multiparty conversations in English, Estonian and Swedish. Each transition was characterised in terms of the number of *intervening intervals* (i.e. silences, overlaps and shorter stretches of solo speech) it took to complete a turn transition. The study found that the distribution of floor transitions was similar to that in Figure 2 with 95% of transitions completed in fewer than 16 intervening intervals. One-interval transitions (i.e. consisting of a single instance of silence or overlap) were the most frequent but they nevertheless accounted for less than 40% of all transitions, suggesting that existing accounts of turn-taking might miss much of floor change dynamics. In addition, transitions involving even numbers of intervals were vanishingly rare, due to the very low likelihood of two or more participants starting or stopping at exactly the same moment.

The composition of transitions in Swedish, Estonian and English in terms of incidence and duration of silent, overlapping and solo-speech intervening overlaps was investigated in Włodarczak and Gilmartin (2021). They found that while one-interval transitions are predominantly silent, more complex patterns of speech and silence were more likely with increasing number of intervening intervals. Overlaps in particular became more common as the number of intervening intervals increased, particularly in BSTs. Similarly, longer transitions were found to involve increasingly many interlocutors speaking, with participation by all three speakers more likely in BST than WST. The authors demonstrated that the most common three-interval transitions (which account for about 21% of transitions identified) were similar across the three data sets, both in terms of interval types and in terms of their percentage frequencies. In other words, even though the transitions are quite complex (especially as the number of intervening intervals increases), a relatively small number of labels accounted for a substantial portion of all floor transitions found. A later study on dyadic phone conversations in the Switchboard corpus found that the transition distribution in Switchboard's conversations broadly followed patterns found in multiparty talk, but that there are fewer complex transitions observed.

3 Method

Below we describe the data used, segmentation and processing into floor state transitions.

3.1 Data

The present investigation is based on a subset of the Berlin Dialogue Corpus (BeDiaCo), version 2 (Belz et al., 2021). The material consisted of free talk and task-oriented interactions between 10 pairs

of German native speakers (mean age = 25.7, SD = 3.8, 10 females, 10 males) in two conditions: face-to-face and remote (Zoom-mediated) conversations. Each of the speaker pairs was living together at the time of the recording.

The conversations were recorded in the phonetics laboratory of the Humboldt Universität zu Berlin. In the face-to-face condition, participants sat opposite each other in a sound-attenuated booth and wore neckband headsets (Beyerdynamics Opus 54) to record their speech. In the remote condition, they were located in adjacent offices and spoke to each other via Zoom installed on two tablets (Lenovo; 10.1 inch). Both tablets were connected to the Internet through the university's wireless network (Eduroam). Subjects wore headphones to listen to each other and their speech was recorded by additional microphones placed in the room (Sennheiser Me62, Sennheiser Me64).

The free conversation had participants talking about self-selected topics (e.g., one's favourite place in Berlin, plans for the next holiday) for about 10 minutes. During the task-oriented part, the speakers participated in the Diapix task. The participants were given about 10 minutes to locate 10–13 differences between their pictures.

The participants came to the lab to be recorded twice, with about a week between sessions. In each session, participants solved two Diapix tasks with a free conversation in between via one medium. For each session and speaker pair, the order of conversation media and Diapix tasks was randomised.

According to the post-experiment questionnaire, 13 of the 20 participants reported using Zoom on a "daily" or "weekly" basis, the others "monthly" or "never". 15 of the participants were "comfortable" or "very comfortable" engaging in Zoom interactions, while five were "neither comfortable nor uncomfortable" (Belz et al., 2021).

3.2 Processing - Identifying speaker transitions

Intervals of speech and silence in each speaker's recording were reconstructed from manually corrected word alignments distributed with the corpus, assembled into talkspurts (or interpausal units, IPUs), given a minimum silence threshold of 200 ms.

The resulting talkspurt segmentation was then used to identify floor state intervals, i.e. divide the conversation into continuous segments where a particular subset of speakers is active. Possible floor states include solo speech by one speaker, intervals of overlapping speech by two speakers, or general silence. More generally, for a conversation with n speakers, there are 2^n possible floor state labels - general silence, n different solo speech labels, and various combinations of speakers in overlap.

In the next step, speaker transitions were identified by locating instances of solo speech of at least one second in duration and recording the floor state intervals between those. Each transition was classified as WST or BST and was characterised by the number of intervening intervals it contained.

4 Results

The corpus comprised 8451 floor transitions (floors defined as talkspurts longer than one second) across 60 conversations. As shown in Figure 3, single-speaker floor constitutes the majority of the data, accounting for 69.2% of the conversation time, followed by silent intervals (23.7%) and overlapping speech by two speakers (7.09%). 3520 between-speaker and 4931 within-speaker transitions were found.

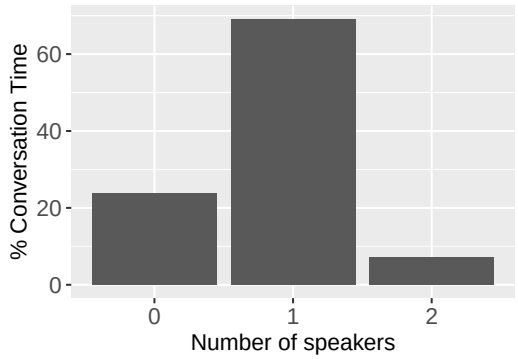

Figure 3: Distribution of the conversation time by the number of speakers.

4.1 General transition patterns

Figure 4 illustrates the percentage of different numbers of between- and within-speaker intervals in Diapix and free conversations in face-to-face (ftf) and Zoom interactions. All groups have more than 98% of transitions completed in less than 15 intervening intervals (Diapix_ftf: 98.25%, Diapix_zoom: 98.26%, free_ftf: 98.50%, free_zoom: 99.34%). In general, the greater the number of intervening intervals involved in the transition, the less frequent they are in the data. For a given number of intervening intervals, there are usually more instances

Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue, September, 11–12, 2024, Trento, Italy.

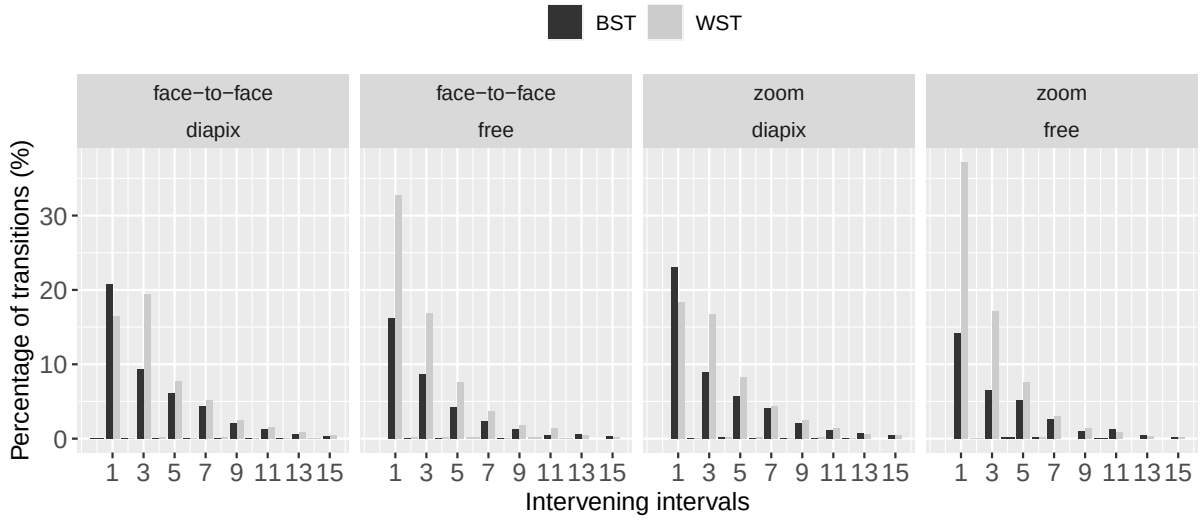

Figure 4: Frequency distribution of speaker transitions in face-to-face and Zoom interactions depending on the number of intervening intervals.

of WST than BST within the group.

Transitions including even numbers of intervening intervals constitute only 0.01% of the data. Such transitions entail two speakers starting or stopping at exactly the same time, with zero gaps and zero overlaps in transition, which is extremely unlikely given the granularity of the manually corrected IPU segmentation.

The cumulative distribution of transitions completed in fewer than 15 intervening intervals is shown in Figure 5. Notably, the difference between the cumulative percentages within each group with the same number of intervening intervals is greater when broken down by task (left panel) than by medium (right panel). Transitions with one intervening interval account for 50.02% of all transitions in free conversation, much higher than the 39.11% in Diapix tasks. Transitions with three to seven intervening intervals exhibit a similar tendency toward a cumulative percentage higher by some distance in free conversations than Diapix. No big differences are found in the cumulative distribution divided by medium, for example, 41.08% for one-interval transitions in face-to-face interactions and 44.55% for Zoom.

In total, 58.35% of transitions are WST. Only 14 conversations have a BST-to-WST ratio above 1, all from Diapix tasks (Figure 6). Compared to free conversations, Diapix tasks have a significantly higher proportion of BST, indicating the Diapix conversations are indeed more interactive and characterised by more frequent speaker change.

Given that the main differences between the media involve floor transitions with one and three intervening intervals, we focus on these to further elucidate the underlying effects of task and medium. Jointly, these cases account for 68.77% of all transitions in the data.

4.2 Transitions with one intervening interval

Unlike the face-to-face and Zoom conversations, which exhibit a similar distribution of intervening intervals per speaker transition, the two tasks show notable differences with respect to transitions consisting of one and three intervening intervals. In the Diapix task, the most common sequence overall was BST with one intervening interval, while WSTs containing one intervening interval were clearly the most frequent sequence in free conversation. In sum, transitions containing only one intervening interval constitute about half of all transition types in free conversations (ftf: 48.50%, zoom: 51.27%), with a slightly lower proportion in Diapix (ftf: 37.01%, zoom: 41.00%).

In order to further elucidate these differences, Figure 7 shows the distribution of all BSTs and WSTs with one intervening interval. In both face-to-face and Zoom interactions, Diapix tasks have a higher proportion of *A_X_B* sequence (between-speaker silences) than *A_X_A* sequence (within-speaker silences); conversely, free conversation shows the opposite pattern. Conversation medium does not seem to affect the frequency of one-interval transitions.

Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue, September, 11–12, 2024, Trento, Italy.

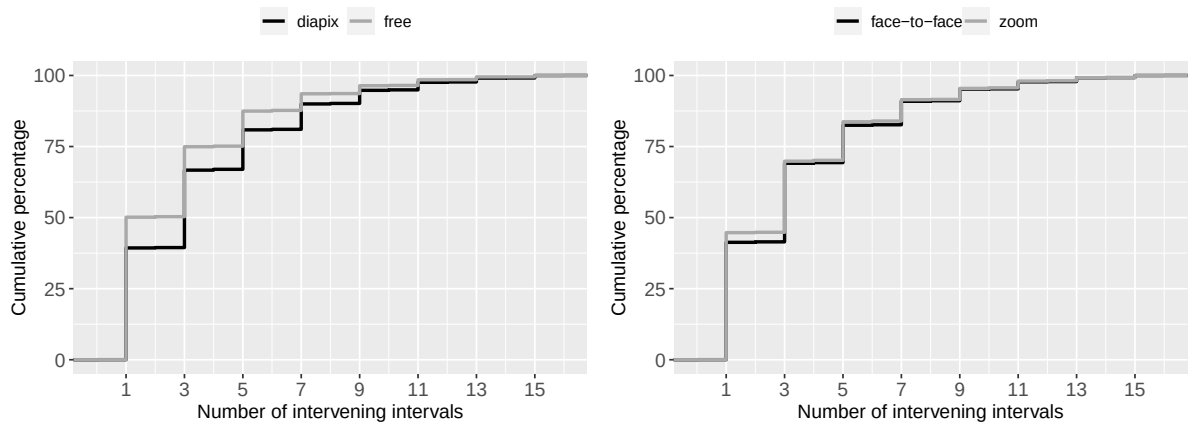

Figure 5: Cumulative distribution of the number of intervening intervals in a speaker transition depending on task (left) and medium (right).

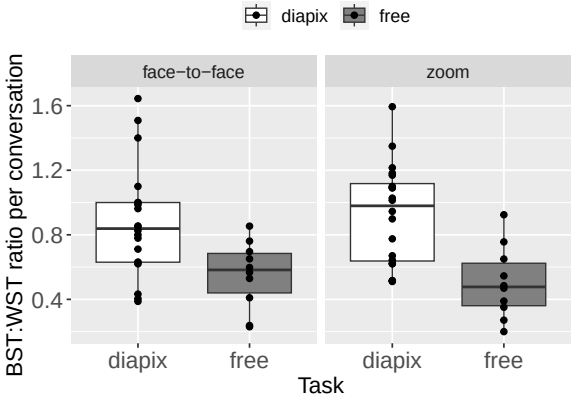

Figure 6: Distribution of BST:WST ratio per conversation.

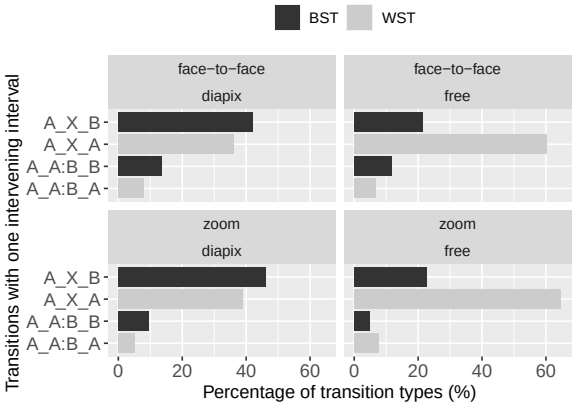

Figure 7: Distribution of floor state sequences in transitions with one intervening interval in face-to-face and Zoom interactions.

4.3 Transitions with three intervening intervals

Overall, transitions containing three intervals account for approximately 25% of all transitions across tasks and media (Diapix_ftf: 28.47%, free_ftf: 25.27%, Diapix_zoom: 25.48%, free_zoom: 23.60%), a slightly higher proportion in Diapix tasks in both media.

Compared to transitions with one intervening interval, there are usually fewer transitions with three intervening intervals across the tasks and media (see Figure 4). Only in the Diapix face-to-face interactions are WSTs containing three intervening intervals more frequent than WSTs containing one intervening interval. However, this difference is not present in Zoom interactions.

In Figure 8, we examine the BSTs and WSTs containing three intervening intervals in more detail. The most frequent transition types are similar for each task, with smaller differences between the media: the most common three-interval sequence used in Diapix conversations is the WST *A_X_B_X_A*, followed by its BST counterpart *A_X_B_X_B* (for an example, see Excerpt 1); while free conversations have a stronger preference for *A_X_A_X_A* sequence, followed by *A_X_B_X_A* in face-to-face interactions and *A_X_A_A:B_A* in Zoom (see Excerpts 2 and 3).

Excerpt 1: Sequences of *A_X_B_X_A* (line 1-3) and *A_X_B_X_B* (line 3-4).

```
1 A: ach so ja aber da sind drei runtergefallen
2 B: nein
3 A: und es hat ein ROTes Rad die Schubkarre
4 B: ja (0.4) und dahinter sind so zwei Stöcker

  A: oh yes, but three of them fell down
  B: no
```


Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue, September, 11–12, 2024, Trento, Italy.

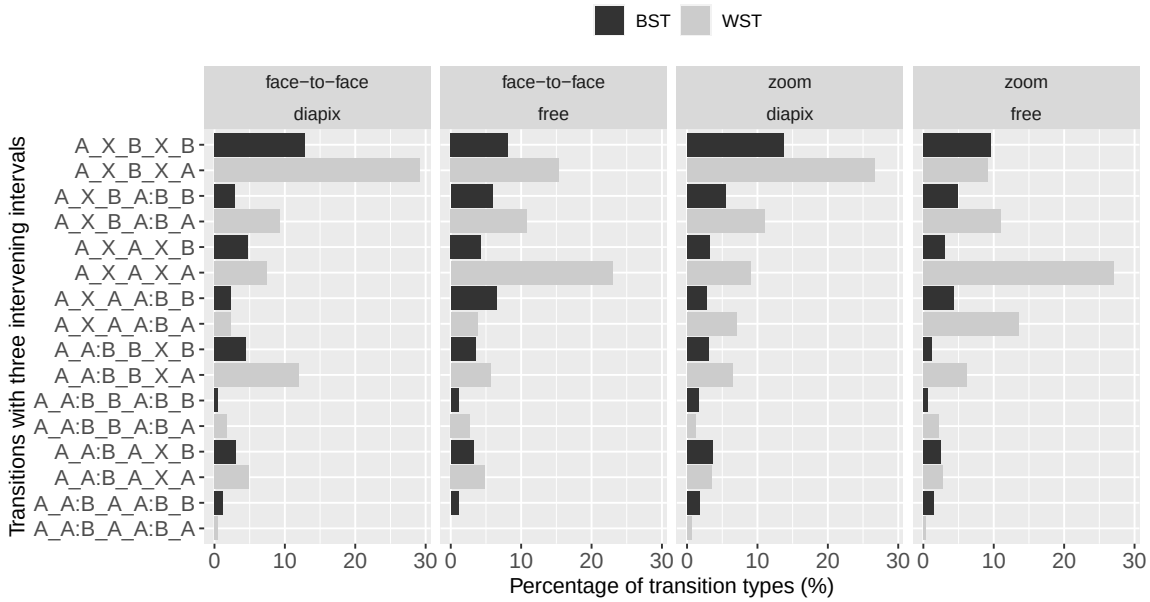

Figure 8: Distribution of floor state sequences in transitions with three intervening intervals in face-to-face and Zoom interactions.

```
A: and it has a RED wheel the wheelbarrow
B: yeah (0.4) and behind it are two sticks
                        [ba_z_diapix2_f2f1:315-323]
```

Excerpt 2: Sequence of *A_X_A_X_A*.

```
1 A: also ich find keine Ahnung mein mein
2    Lieblingsort in Berlin sind so (1.2) Orte zu
3    denen man sehr oft eigentlich hingeht (1.4)
4    Mercedes Benz Arena is für MICH voll schön

  A: well I don't know, my favorite places in
     Berlin are (1.2) actually places that you
     frequently visit (1.4) for ME Mercedes Benz
     Arena is quite nice
                        [bd_z_frei_m8f7:337-350]
```

Excerpt 3: Sequence of *A_X_A_A:B_A*.

```
1 A: diese Frage ganz anders beantworten (0.2)
2    auf seine eigene [Art und Wei]se
3 B:                  [ja=        ]

  A: answer this question quite differently (0.2)
     in his [own way]
  B:        [yeah   ]
                        [bd_z_frei_m8f7:502-506]
```

Upon closer examination of WSTs containing three intervening intervals, as shown in Figure 9, we can see that these sequences appear to fall into two distinct groups, depending on whether the interlocutor B is involved during the transitions. Transition including the involvement of the other interlocutor is preferred in all groups. Yet, transitions without B's involvement constitute a higher percentage in free conversations, regardless of medium. The implication of these results will be discussed in the next section.

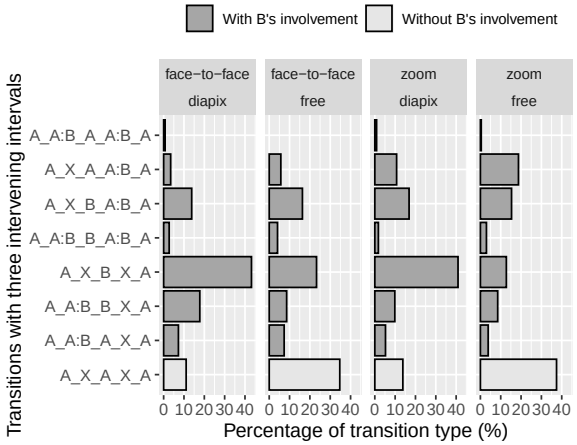

Figure 9: Distribution of WSTs with three intervening intervals categorised by B's involvement.

5 Discussion

The results of the present study show a striking similarity to previous studies (Gilmartin et al., 2020; Gilmartin, 2021), especially the transition pattern in free conversations. First, the distribution frequency of BSTs and WSTs declines sharply when the number of intervening intervals increases. Second, there are always more WSTs than BSTs within the group with the same number of intervening intervals.

However, Diapix task exhibits a distinct feature in both BSTs and WSTs containing one and three intervening intervals. Among the one-interval tran-

Proceedings of the 28th Workshop on the Semantics and Pragmatics of Dialogue, September, 11–12, 2024, Trento, Italy.

sitions, the BST occurrences outnumbered those of WST. To be more specific, there are more *A_X_B* sequences compared to the *A_X_A* sequences. This difference indicates that Diapix conversations are indeed characterised by high interactivity, with interlocutors changing the floor more frequently to facilitate intensive information exchange. Rapid turn-taking leads to a high number of between-speaker gaps, while monologic utterances are marked by numerous within-speaker pauses.

In the case of WSTs containing three intervening intervals, their majority consists of transitions where B produced a short utterance during A's monologic stretch, see Figure 9. The results suggest that Diapix tasks prompt speakers to provide more short utterances (e.g. backchannel, acknowledgement) than free conversations. These findings highlight that task-based conversations exhibit different turn-taking dynamics compared to free conversations. Our results thus reflect the distinctive characteristics of conversations with different levels of interactivity.

Compared to the task, the medium seems to play a less important role in speaker transition patterns. We expected that online-mediated conversations would reduce interlocutors' engagement, resulting in fewer speaker changes. A noteworthy difference is observed in the three-interval WSTs in Diapix tasks, where their occurrences surpassed those of one-interval WSTs in face-to-face interactions, but not over Zoom. We assume that interlocutors provide more feedback in the back channel when conversing face-to-face, while on Zoom, due to the latency and remoteness, backchannel-like utterances are avoided to prevent misinterpretation as a turn-starter, which could cause unintended interruption.

Beyond this, the media did not alter the general speaker transition patterns within the same task. This may be attributed to interlocutors' increased familiarity with remote conversations (see Section 3.1), leading them to adapt to the new conversation dynamics.

Another potential explanation for the minor influence of remote conversation is that the transition delay of audio signal does not reach the threshold needed to create noticeable disruptions, such as the 800 ms delay in telephone conversation suggested by Egger-Lampl et al. (2010). Unfortunately, we were not able to obtain the exact delay in real-time conversation, as Zoom does not provide access to this data. Consequently, this forms a new area of focus for future work where the latency in remote conversations will be examined.

It is worth pointing out that data-driven analysis such as that described above cannot capture all the details of real conversations. Based on the task settings, we assume that the three-interval WSTs are primarily short feedback utterances, such as acknowledgement, short answers and backchannelling. Nonetheless, instances of unsuccessful floor competition and premature relinquishment would also be included in these sequences. A qualitative analysis of these cases is needed to determine the exact distribution of backchannelling and other potential turn-taking behaviours.

We plan to build on our analyses by exploring the role of the duration of the intervening intervals in transitions and indeed the stretches of solo speech bounding the transitions in order to deepen our understanding of how speech is arranged by participants, and also to extend our analyses to a variety of spoken interaction types. We hope that the insights gained by these studies will contribute to a better understanding of human-human spoken interaction and will aid in specifying more effective artificial dialogue technologies.

Acknowledgments

Emer Gilmartin's work was supported by Institute of Information communications Technology Planning Evaluation (IITP) grant funded by the Korea government(MSIT) (RS-2022-II220043,Adaptive Personality for Intelligent Agents)

References

Jeremy N. Bailenson. 2021. Nonverbal overload: A theoretical argument for the causes of Zoom fatigue. *Technology, Mind, and Behavior*, 2(1).

Rachel Baker and Valerie Hazan. 2011. DiapixUK: Task materials for the elicitation of multiple spontaneous speech dialogs. *Behavior Research Methods*, 43(3):761–770.

Beatrice Beebe, Diane Alson, Joseph Jaffe, Stanley Feldstein, and Cynthia Crown. 1988. Vocal congruence in mother-infant play. *Journal of Psycholinguistic Research*, 17:245–259.

Beatrice Beebe, Joseph Jaffe, Frank Lachmann, Stanley Feldstein, Cynthia Crown, and Michael Jasnow. 2000. Systems models in development and psychoanalysis: The case of vocal rhythm coordination and attachment. *Infant Mental Health Journal*, 21(1-2):99–122.

Malte Belz, Alina Zöllner, Megumi Terada, Robert Lange, Lea-Sophie Adam, and Bianca Sell. 2021. Dokumentation und Annotationsrichtlinien für das Korpus BeDiaCo.

Julie E. Boland, Pedro Fonseca, Ilana Mermelstein, and Myles Williamson. 2021. Zoom disrupts the rhythm of conversation. *Journal of Experimental Psychology: General*, pages 1272–1282.

Paul T. Brady. 1968. A statistical analysis of on-off patterns in 16 conversations. *Bell System Technical Journal*, 47(1):73–91.

Oliveira Maggie Bullock and Bianca Sell. 2022. PDF and PSD files of DiapixGEtv picture materials – German version adapted to elicit tense vowels.

Carole Edelsky. 1981. Who's Got the Floor? *Language in Society*, 10(3):383–421.

Sebastian Egger-Lampl, Raimund Schatz, and Stefan Scherer. 2010. It takes two to tango - Assessing the impact of delay on conversational interactivity on perceived speech quality. In *Proceedings of the 11th Annual Conference of the International Speech Communication Association, INTERSPEECH 2010*, pages 1321–1324.

Emer Gilmartin. 2021. *Composition and Dynamics of Multiparty Casual Conversation: A Corpus-based Analysis*. Ph.D. thesis, Trinity College, Dublin, Dublin, Ireland.

Emer Gilmartin, Kätlin Aare, Maria O'Reilly, and Marcin Wlodarczak. 2020. Between and within speaker transitions in multiparty conversation. In *Proceedings of Speech Prosody 2020*, pages 799–803, Tokyo, Japan.

Leilani H. Gilpin, Danielle M. Olson, and Tarfah Alrashed. 2018. Perception of speaker personality traits using speech signals. In *Extended Abstracts of the 2018 CHI Conference on Human Factors in Computing Systems*, page LBW514. ACM.

Mattias Heldner and Jens Edlund. 2010. Pauses, gaps and overlaps in conversations. *Journal of Phonetics*, 38(4):555–568.

Joseph Jaffe, Louis Cassotta, and Stanley Feldstein. 1964. Markovian model of time patterns of speech. *Science*, 144(3620):884–886.

Joseph Jaffe and Stanley Feldstein. 1970. *Rhythms of dialogue*. Academic Press, New York.

Kornel Laskowski. 2011. *Predicting, detecting and explaining the occurrence of vocal activity in multiparty conversation*. Ph.D. thesis, Carnegie Mellon University.

A. C. Norwine and O. J. Murphy. 1938. Characteristic time intervals in telephonic conversation. *Bell System Technical Journal*, 17(2):281–291.

Daniel C. O'Connell, Sabine Kowal, and Erika Kaltenbacher. 1990. Turn-taking: A critical analysis of the research tradition. *Journal of psycholinguistic research*, 19(6):345–373.

Harvey Sacks, Emanuel A. Schegloff, and Gail Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Langauge*, 50(4):696–735.

Deborah Tannen. 1980. Toward a Theory of Conversational Style: The Machine-Gun Question. Technical report, Southwest Educational Development Laboratory, 211 East 7th Street, Austin, Texas 78701.

Louis ten Bosch, Nelleke Oostdijk, and Lou Boves. 2005. On temporal aspects of turn taking in conversational dialogues. *Speech Communication*, 47(1–2):80–86.

Louis ten Bosch, Nelleke Oostdijk, and Jan Peter de Ruiter. 2004. Durational aspects of turn-taking in spontaneous face-to-face and telephone dialogues. In *Proceedings of 7th International Conference on Text, Speech and Dialogue*, Brno, Czech Republic.

Kristin J. Van Engen, Melissa Baese-Berk, Rachel E. Baker, Arim Choi, Midam Kim, and Ann R. Bradlow. 2010. The wildcat corpus of native-and foreign-accented English: Communicative efficiency across conversational dyads with varying language alignment profiles. *Language and Speech*, 53(4):510–540.

Marcin Włodarczak and Emer Gilmartin. 2021. Speaker transition patterns in three-party conversation: Evidence from English, Estonian and Swedish. In *Proceedings of Interspeech 2021*, pages 801–805.