Speaker transitions in 2- and 3-party conversation

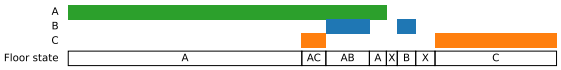
Emer Gilmartin
ADAPT Centre, Trinity College Dublin
Ireland
gilmare@tcd.ie

Marcin Włodarczak
Stockholm University
Sweden
wlodarczak@ling.su.se

1 Introduction

This paper reports ongoing work on temporal aspects of how participants manage conversation. We analyse dyadic phone conversations in the Switchboard (SWB) corpus (Godfrey et al., 1992) using a method previously employed on multiparty dialogue (Włodarczak and Gilmartin, 2021). The analysis is based on *floor state* - who is speaking or silent at any moment during interaction. By annotating *floor state intervals*, stretches of time during which a particular floor state holds, we can analyse *floor state transitions* or sequences of contiguous floor states. We are particularly interested in transitions between 'substantial' stretches of single party speech, to elucidate turntaking. We focus on transitions between stretches of single party speech in the clear of at least one second in duration (to avoid treating e.g. backchannels as turns). We distinguish *between speaker transitions* (BST) and *within speaker transitions* (WST). In WST, the speaker on either side of the transition is the same, as in turn retention, while in BSTs, the single party speech bounding the transition is by different speakers, as in turn change. To illustrate, Figure 1 shows a short exchange from a 3-party conversation. It involves 8 floor states – solo speech (**A**, **B**, **C**), overlaps (**AC**, **AB**) and general silence (**X**). Without the one second threshold we would treat this stretch as a series of three transitions: **A_AC_AB_A** from A to A, **A_X_B** from A to B, and **B_X_C** from B to C. However, looking at the transcript and the speech patterns, it seems more likely that the *longer* stretches of solo speech (**A**, **C**) delimit a single more complex transfer of floor possession from speaker A to C.

In previous work we found similarities in speaker transition distribution in different multiparty corpora. One-interval transitions were the largest class for all corpora studied, with a higher proportion of one-interval transitions in WST. However, less than half of between and within speaker transitions were accomplished with a single intervening interval of silence or overlap, indicating that turn change and retention are often a more complex sequence of events than a simple silence or short overlap. We found high levels of uniformity in the most common WSTs and BSTs found in different languages and settings (Gilmartin et al., 2020, 2019; Gilmartin, 2021; Włodarczak and Gilmartin, 2021; Gilmartin et al., 2018). We found considerable complexity and growing incidence of participation by more speakers with transition length, and that silent intervals account for a significant part of transition duration. Below, we analyse SWB to investigate whether our findings on multiparty talk hold for dyadic phone conversations.



A …cameras. You're just not normally as aware of it.
C Yeah
B It's true.
B Yeah.
C There's two on every bus.

Figure 1: An excerpt from a 3-party casual conversation corresponding to a between-speaker transition, **A_AC_AB_A_X_B_X_C**, from speaker A to C with six intervening intervals (**AC**, **AB**, **A**, **X**, **B**, **X**). *Top*: Temporal organization of individual speakers' contributions (represented as color bars) and the resulting floor states. *Bottom*: Simplified transcript. Speakers' contributions are color-coded for consistency.

2 Data and Annotation

We used the 2438 dyadic phone conversations (259 hours) in the Switchboard-1 Telephone Speech Corpus: Release 2, with the Mississippi University ISIP word level transcription. Transcripts were processed using Praat and Python to create speech and silence labels with all non-speech sounds suppressed to silence, resulting in 520135 talkspurts.

Proceedings of the 26th Workshop on the Semantics and Pragmatics of Dialogue,
August, 22-24, 2022, Dublin.

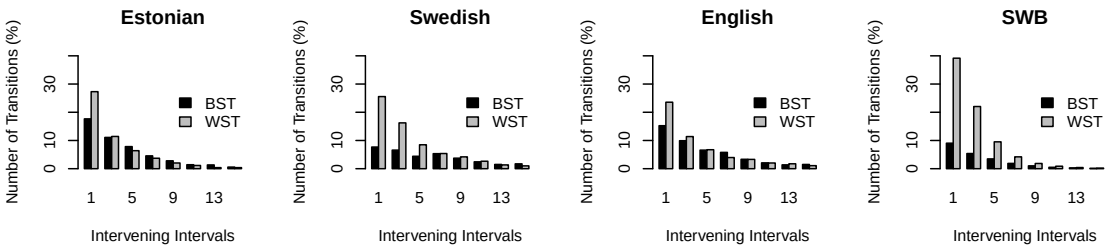


Figure 2: Distribution of Between and Within Speaker Transitions in Switchboard and 3 other corpora

From these we generated BST and WST labels, as described above. We used annotations from 3-party dialogue data from our previous studies to compare with the SWB results –three-party spontaneous conversations in Estonian (Lippus et al.) and Swedish (Włodarczak and Heldner, 2017), and collaborative conversational games in English (Litman et al., 2016). This data set contained 22106 talkspurts in 9 hours and 51 minutes hours of conversation.

3 Results

Results are first presented for SWB, and then contrasted with results on multiparty corpora.

Distribution of Speech, Silence and Overlap SWB has lower incidence of silence and overlap than the multiparty datasets, and higher incidence of single-party speech in the clear.

Distribution of speaker transitions SWB yielded 256,655 speaker transitions in 259 hours of talk, an average of one every 3.7 seconds. In the 3-party data, there was an average of one transition every 4.7 seconds. The vast bulk (over 99%) of transitions in SWB comprised fewer than 16 intervening intervals (approximately 99%). There were vanishingly few transitions involving even numbers of intervening intervals (47 out of 256,655). One-interval transitions are the largest class, and the frequency of transitions decreases with increasing numbers of intervening intervals. All of these results reflected our earlier findings for 3-party data.

Distribution of BSTs and WSTs In SWB, 78.28% of transitions are WST, greatly outnumbering BSTs. WSTs account for 81% of 1-interval transitions, 80% of 3-interval, with proportion falling with increasing numbers of intervals to 60% of 15 interval transitions. Figure 2 shows the split between BSTs and WSTs for odd number interval transitions in SWB and in the 3-party conversations. In SWB, 47.72% of all transitions (41.65% of BSTs and 50.03% of WSTs) were accomplished with one intervening interval, 27.14% (24.77% of BSTs and 28.15% of WSTs) with two intervening intervals, and 12.86% (15.98% of BSTs and 12.16% of WSTs) with 3 intervening intervals

4 Discussion

SWB has less silence and overlap and more speech in the clear than the 3-party data - this may be due to modality; on the phone, speakers may wait for their interlocutor to finish before commencing to speak, and may give less verbal feedback in overlap. It could also reflect differences between dyadic and multi-party talk. The distribution of speaker transitions largely reflects results from the 3-party data (and also from 4- and 5- party data analysed in (Gilmartin, 2021). The largest category are 1-interval transitions, even-number interval transitions are extremely rare, and the number of transitions drops off with increasing numbers of intervals. The proportion of 1-interval transitions in SWB is greater than in 3-party, but still only accounts for 47.7% of all transitions, highlighting how most transitions involve more than a single silence or overlap, even in dyadic phone conversations. The higher incidence of WSTs than BSTs in SWB reflects results in the 3-party data. WSTs more dramatically outnumber BSTs in SWB than in the 3-party data. This could reflect long turns being taken in SWB, perhaps because participants were strangers, or indeed, may be a feature of telephone conversation.

Our analysis has shown that more than half of all BSTs and WSTs involve more than one intervening interval of speech, silence or overlap between longer stretches of single party speech. This reflects previous results on multiparty spoken interaction, implying that turn change and retention even in dyadic phone conversations exhibit a level of complexity that is not covered by modelling them as a simple gap or overlap.

5 Acknowledgements

This work was conducted with the support of Science Foundation Ireland under Grant Agreement No. 13/RC/2106 at the ADAPT SFI Research Centre at Trinity College Dublin. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant Number 13/RC/2106. The work was also funded by Swedish Research Council project 2019-02932 *Prosodic functions of voice quality dynamics* to Marcin Włodarczak.

References

E. Gilmartin, Christian Saam, Carl Vogel, Nick Campbell, and Vincent Wade. 2018. Just talking - modelling casual conversation. In *Proceedings SIGdial 2018*, pages 51–59, Melbourne, Australia.

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

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.

Emer Gilmartin, Mingzhi Yu, and Diane Litman. 2019. Comparing speech, silence and overlap dynamics in a task-based game and casual conversation. In *Proceedings of ICPhS 2019*, pages 3408–3412.

John J. Godfrey, Edward C. Holliman, and Jane McDaniel. 1992. SWITCHBOARD: Telephone speech corpus for research and development. In *Acoustics, Speech, and Signal Processing, 1992. ICASSP-92., 1992 IEEE International Conference on*, pages 517–520.

Pärtel Lippus, Tuuli Tuisk, Nele Salvestre, and Pire Tiras. Phonetic corpus of Estonian spontaneous speech.

Diane Litman, Susannah Paletz, Zahra Rahimi, Stefani Allegretti, and Caitlin Rice. 2016. The teams corpus and entrainment in multi-party spoken dialogues. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 1421–1431.

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.

Marcin Włodarczak and Mattias Heldner. 2017. Respiratory constraints in verbal and non-verbal communication. *Frontiers in Psychology*, 8:708.