Getting from A to B: Exploring Floor State Transitions in Conversation

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

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

1 Introduction

In this abstract we describe ongoing analysis of multiparty spoken interaction, where participants start and stop speaking, taking and relinquishing the floor and locally arranging turn change and retention (Sacks et al., 1974). We consider conversation in terms of stretches of speech and silence, using only timing information for analysis. We have devised a method of labelling to capture *floor state transitions* - sequences of speech and silence involved in transitions from a stretch of one party speech (speech in the clear) by one speaker to the next stretch of one party speech by the same speaker (within speaker transition - WST) or another speaker (between speaker transition - BST). To approximate turn changes and retention, we impose left and right hand side minimum duration thresholds on the single party speech in the clear bordering the transitions. We dub the intervals of speech, silence and overlap between the single-speaker stretches *intervening* intervals. We have been analysing patterns of intervening intervals in multiparty talk, concentrating on 3-party interaction in Estonian, Swedish, and English. Below we briefly explain the labelling scheme and summarize results to date in this work.

2 Labelling Scheme

We define the 'floor state' at any point of a conversation as the totality of participants speaking, and represent interaction as a series of labels for intervals of varying where a particular floor state prevails (see Figure 1). For example, an interval where A and B are speaking in overlap is labelled AB, C speaking alone is labelled C, and general silence X. In n-party speech, there are 2^n possible floor states, so 3-party speech could include any of the 8 labels: X, A, B, C, AB, AC, BC, ABC. We define a transition as the set of intervening intervals between two stretches of single party speech. We impose left and right hand 1-second minimum thresholds on the single party speech, generating *1Sp1-1Sp1* transitions, which can be BST or WST.

Figure 1 shows a stretch of talk with three instances of two-party overlap (AB, AC, AC), an instance a three-party overlap (ABC), and three intervals of solo speech (A, B, C). We can define a five-interval BST from A to C comprising **AB_ABC_AB_B_BC**. Note that if the right hand one-second threshold were not applied, the example would be classified as involving two transitions (from A to B and from B to C), even though the short stretch of solo speech by B is unlikely to be a claim for turn possession.

We process segmentation data from spoken interaction with a Python script using TextGridTools (Buschmeier and Włodarczak, 2013) to create floor state and transition labels and extract the number and identity of participants speaking during the transition. All code and annotation are available at https://zenodo.org/record/4923246.

3 Summary of Results to Date

We have used these labels to explore floor state transitions in corpora of three-party spontaneous conversations in Estonian (Lippus et al.), Swedish (Włodarczak and Heldner, 2017), and the TEAMS corpus of collaborative conversational games in English (Litman et al., 2016), and also in casual multiparty talk in English.

In all cases, the majority of transitions involve more than one intervening interval to complete and the vast bulk of transitions involve odd numbers of intervening intervals (Gilmartin et al., 2019, 2020; Włodarczak and Gilmartin, to appear). The scarcity of even numbers of intervening intervals follows from the rarity of smooth switches and instances of simultaneous onset or offset of speech. We have

Proceedings of the 25th Workshop on the Semantics and Pragmatics of Dialogue, September 20–22, 2021, Potsdam / The Internet.

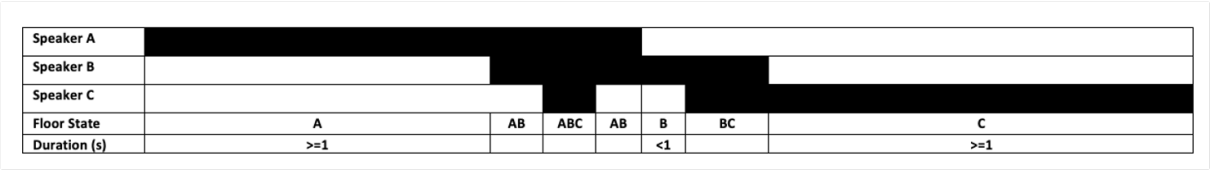

Figure 1: An example of a between-speaker transition.

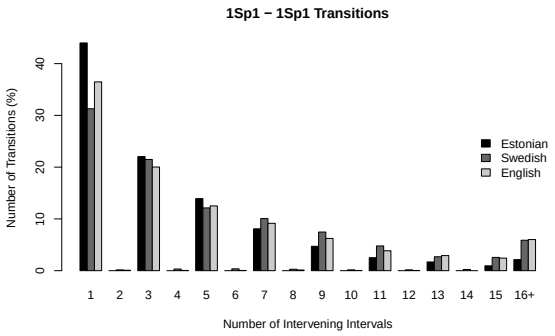

Figure 2: Number of floor state intervals between single-speaker intervals of 1 second or more in duration - Estonian, Swedish, and English 3-party conversation

found that one-interval transitions are the largest class, and the frequency of transitions decreases with increasing numbers of intervening intervals, as shown in Figure 2. We have also found that WSTs account for more of the one-interval transitions, particularly around silence, perhaps due to breathing pauses.

In terms of speaker participation in transitions, one-interval transitions are silence or overlap,with 0 or 2 speakers involved. With more intervals, transitions can have more participants, with participation by all three speakers more likely in BST than WST. In the Estonian, Swedish, and Teams data, silence was present in over 90% of transitions, with overlap appearing in 53%. With more intervals, transitions contain more complex combinations of speech and silence, and all of these features become more likely. Silence occurs in the vast bulk of transitions of 3 or more intervals, as does solo speech. The incidence of overlap increases with increasing number of intervals, and is more common throughout in BST than WST. (Wlodarczak and Gilmartin, to appear)

We found that silence accounts for a large share of the duration of 3-intervals WSTs and BSTs, and remains the lead component in terms of duration but decreases with increasing numbers of intervening intervals, while the duration from overlap increases (Wlodarczak and Gilmartin, to appear).

The distribution of the most common transition sequences across the datasets are similar. In all cases, the most common sequences overall were A_X_A (within speaker silence) and A_X_B (between speaker silence). Interestingly, for WST, both A_X_A_X_A and A_X_B_X_A were more common than 1-interval overlap (A_A:B_A), while the second most common BST was 1-interval overlap (A_A:B_B).

Almost 60% of all transitions are 1- or 3-interval. For all languages, the most common 5-interval transitions were less frequent than the fifth most frequent 3-intervals, except for English WSTs, where the most common 5-interval transition was marginally more frequent than the fifth most frequent 3-interval WST. We therefore further explored the 3-interval transitions to understand the most common transitions in the data. The first and second most common 3-interval WST and BST sequences for were the same in all three languages analysed. All of the top five 3-interval within speaker transitions across the three languages are accounted for by six transition labels, while the top five between speaker transitions are covered by seven transition labels. The categories also show great similarity in their percentage frequencies across the three datasets.

4 Ongoing Work

Our explorations have shown interesting results on the composition of within and between speaker transitions in multiparty talk, with similarities in how these occur across the languages we have analysed. We are expanding the analysis to other corpora, including dyadic speech, to see how well our findings generalize. We intend to create an inventory of transitions most commonly found across a large number of corpora, and will perform detailed phonetic analysis of the more common sequences. This will add to our understanding of how spoken interaction works, as well as inform design of more appropriate spoken dialog technology in applications requiring human like behaviour.

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

Hendrik Buschmeier and Marcin Włodarczak. 2013. TextGridTools: A TextGrid processing and analysis toolkit for Python. In *Tagungsband der 24. Konferenz zur Elektronischen Sprachsignalverarbeitung (ESSV 2013)*, volume 65 of *Studientexte zur Sprachkommunikation*, pages 152–157, Dresden. TUDpress.

E. Gilmartin, Kätlin Aare, M. O'Reilly, and M. Wlodarczak. 2020. Between and within speaker transitions in multiparty conversation. In *Speech Prosody*, pages 799–803.

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.

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.

H. Sacks, E.A. Schegloff, and G. Jefferson. 1974. A simplest systematics for the organization of turn-taking for conversation. *Language*, pages 696–735.

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

Martin Wlodarczak and Emer Gilmartin. to appear. Speaker transition patterns in three-party conversation:
evidence from english, estonian and swedish. In *Interspeech 2021, 22nd Annual Conference of the International Speech Communication Association, Brno, Czechia, 30 August - 3 September 2021*. ISCA.