Originally published in: Proceedings of the FONETIK (Swedish Phonetics Conference) 2019 in Stockholm, June 10–12, 2019. - Stockholm: Stockholm University, 2019. Pp. 119-124

Timing properties of hand gestures and their lexical counterparts at turn transition places

Margaret Zellers[1], Jan Gorisch[2], David House[3], Benno Peters[1]
[1] Institute for Scandinavian Studies, Frisian Studies, and General Linguistics, University of Kiel, Germany
[2] Leibniz-Institute for the German Language, Mannheim, Germany
[3] Speech, Music & Hearing, KTH Royal Institute of Technology, Stockholm, Sweden
mzellers@isfas.uni-kiel.de, gorisch@ids-mannheim.de,
davidh@kth.se, peters@ipds.uni-kiel.de

Abstract

Looking at gestures as a means for communication, they can serve conversational participants at several levels. As co-speech gestures, they can add information to the verbally expressed content and they can serve to manage turn-taking. In order to look closer at the interplay between these resources in face-to-face conversation, we annotated hand gestures, syntactic completion points and the related turn-organisation, and measured the timing of gesture strokes and their lexical/phrasal referent. In a case study on German, we observe the trend that speakers vary less in gesture-lexis on- and offsets when keeping the turn after syntactic completions than at speaker changes, backchannel or other locations of a conversation. This indicates that timing properties of non-verbal cues interact with verbal cues to manage turn-taking.

Introduction

Everyday conversation, the fundamental context in which spoken language is used, has been demonstrated to have consistent structural features to which conversational participants orient, in particular with regard to turn-taking. Sacks et al. (1974) report that at Transition Relevance Places—i.e. locations where speaker change may become relevant—new speakers have priority to take up a turn, with the current speaker only continuing if a new speaker does not take the floor. Thus, the current speaker must have ways of communicating her/his intention to hold or cede the floor to an interlocutor (or interlocutors).

A variety of communicative means have been proposed by which floor-holding and floor-ceding can be achieved in conversation. These can be broadly grouped into the categories of linguistic, phonetic, and gestural means. By *linguistic* means, we primarily refer to syntactic or semantic completion of an utterance in context. *Phonetic* means may include such features as pitch variation (choice of contour or size of pitch movements), amplitude variation, and speech rate variation. *Gestural* means can include body movements of any type, such as those of the eyes, eyebrows, head, and/or hands. A large body of literature has investigated the role and interplay of linguistic and phonetic cues at turn boundaries, suggesting that syntactic/semantic completion is a strong cue to speaker change, while pitch, phonation quality, and duration variation can also contribute as turn-taking cues (Schaffer, 1983; Auer, 1996; Local, Kelly, and Wells, 1986; Koiso, Horiuchi, Tutiya, Ichikawa, and Den, 1998; Gravano and Hirschberg, 2009, 2011; Kane, Yanushevskaya, de Looze, Vaughan, and Ní Chasaide, 2014; Heldner and Włodarczak, 2015; Zellers, 2017, *inter alia*). Similarly, a variety of gestural cues have been shown to impact turn-taking, including gaze direction (Edlund and Beskow, 2007, 2009) and hand movements (Streeck and Hartge, 1992; Mondada, 2007; Sikveland and Ogden, 2012).

Since some aspects of turn-taking signalling involve the linguistic system, it is particularly interesting to make cross-linguistic comparisons. We report data from a larger project investigating gesture and phonetic features at turn boundaries in Swedish and German.

Gesture and turn-taking

A previous study (Zellers et al., submitted) investigated which phase hand gestures were in (i.e. preparation, hold, stroke, retraction, cf. Kendon, 2004) at the time that speech ended. The current work makes use of the data and annotations from the previous study.

Material and methods

Our Swedish data consists of five five-minute chunks of conversations from Spontal (Edlund et al., 2010), comprising ten participants in total (8 male, 2 female). The German data come from three 7-minute chunks of conversation taken from the FOLK corpus (Schmidt, 2014), comprising 4 participants (all male). While we attempt to make cross-linguistic comparisons, these comparisons are also mediated by the differences in interactional setting in the two corpora, which are an unavoidable artefact of the existing available data.

We annotated syntactic/semantic completion with reference to the orthographic transcription. Phonetic annotations were carried out in Praat (Boersma & Weenink, 2018) without access to the video signal. Conversely, gestures were annotated using ELAN (Version 5.4, Max Planck Institute, 2019) without using the audio signal. We also annotated the data for the type of transition between the current and the next turn (whether within the same speaker or different speakers).

Transition types

Of particular interest to us are places in conversation where speaker change could become relevant. These locations were defined using two criteria: first, the presence of a silent pause, and second, the potential syntactic/semantic completion in context of the lexical material at that location. Locations meeting these criteria were given a label defining the turn-taking behaviour at that point:

- *Change*: the current speaker produces a complete full turn in declarative form, and then the next speaker launches a full turn
- *Keep*: the current speaker produces a complete full turn in declarative form, and then the same speaker launches an additional full turn
- *Backchannel*: the current speaker produces a complete full turn in declarative form, the other speaker produces a short response token (e.g. *ja, mhm*), and then the first speaker launches an additional full turn
- *Question*: the current speaker produces a complete full turn with lexical/syntactic interrogative form, and then the next speaker launches a full turn

Results

For the 98 (German) and 102 (Swedish) cases where hand gesture occurred in the vicinity of the offset of speech, the distribution of gesture phases is shown in Figure 1. Ongoing gestures of all kinds at the offset of speech were much more frequent at Backchannel and Keep locations than at Changes and Questions in both languages. In terms of gesture phases, gesture strokes co-occurring with the offset of speech only occur at Keep and Backchannel locations, while the other gesture phases can occur at Backchannels, Changes, and Keeps.

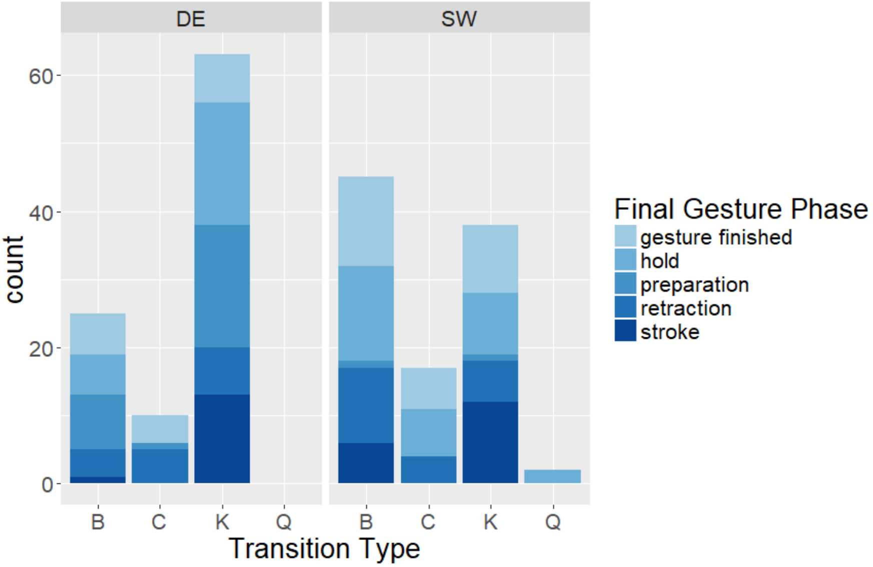

Figure 1. Gesture phase at offset of speech in German (DE) and Swedish (SW), from Zellers et al., submitted. Transition types are B(ackchannel), C(hange), K(eep), and Q(uestion).

Gestures and referentiality: a case study

Many studies of gesture distinguish between gestures that contain some kind of semantic material (*referential* gestures) and those which are primarily rhythmic in nature (*beat* gestures) (cf. e.g. McNeill & Levy, 1982; Prieto et al., 2018). Since referential gestures contain semantic material which is presumably absent from beat gestures, it is possible that they may make different contributions to turn-taking in conversation as well.

Material and methods

Using a subset of the German data from our previous study, we labelled the gesture strokes for whether they were referential (i.e. was there a spoken lexical item with which they plausibly shared semantic material) or not. We also labelled the location of the lexical item (a word or a short phrase) with which the referential gesture shared its semantic material. Using a Praat script, we extracted the locations of the gestures and the lexical items, as well as the turn-taking features labelled for those locations. The current study includes not only gestures at potential turn boundaries, but also those which are turn-internal (labelled NONE in the following). Thus a total of 147 referential gesture strokes and their corresponding lexical items from two speakers are investigated.

Results

Since only two speakers are included in the current study, it was first important to determine whether the speakers' alignment of referential gestures and their corresponding lexical items was similar or not. T-tests showed that while both speakers tended to start their gesture strokes about 0.106 s before the related lexical item, speaker TB tended to end his gestures about 0.09 s earlier than speaker TN, although this difference did not quite attain statistical significance (t = -1.848, df = 105.18, p = .067). No significant differences in overall gesture duration could be found between the speakers.

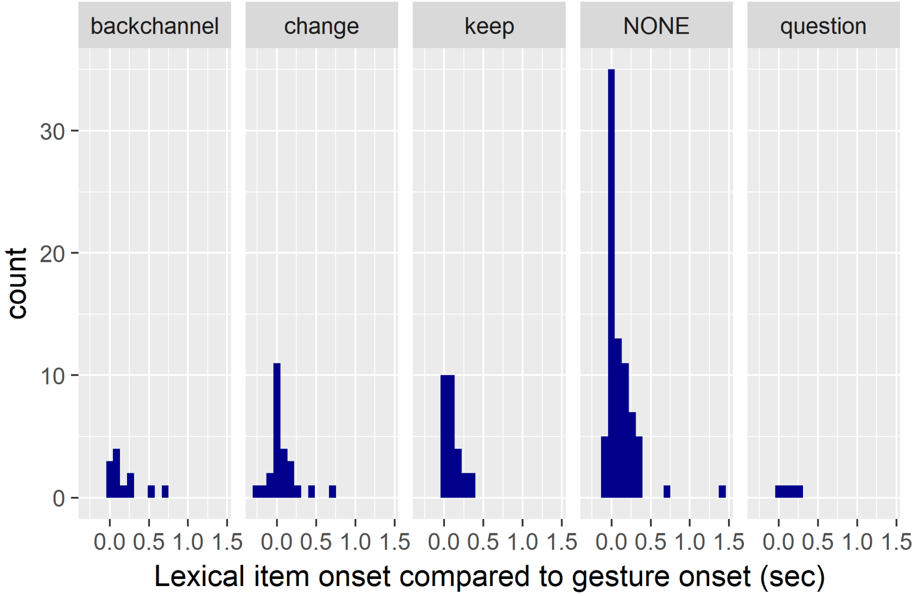

Figure 2. Distribution of lexical item onset compared to gesture stroke onset at different possible transition locations and turn-internally (NONE). 0 on the x-axis is equivalent to the gesture stroke onset.

Although the overall tendency was for speakers to begin their referential gesture strokes before the onset of the spoken word with which they were connected, Figure 2 shows that there are some important distributional differences in what kinds of timings were possible. At some turn transition locations, namely Backchannels, Keeps, and Questions, the onset of the lexical word always began simultaneously with or after the hand gesture onset, while in Changes and in turn-internal gestures, the lexical word could begin earlier than the accompanying referential hand gesture. Additionally, the relative timing of referential hand gestures and their accompanying lexical items appears to be rather narrowly constrained in Keeps, while it varies more freely in other contexts (there are only 4 Question items, so the timing distribution here may not be complete).

Discussion

We have thus far only investigated a small subset of the gestures in our data set with reference to the behavior of referential hand gestures and their related lexical items, and thus the current study should be interpreted more along the lines of a case study than as a definitive investigation. However, we identify some interesting trends in our data, which we look forward to investigating in a larger data set and across multiple languages.

Our previous work as well as that of other researchers suggests that Keeps, i.e. contexts in which a current speaker wishes to hold the floor despite the default possibility of another speaker taking up a new turn, are locations where speakers must make a particular effort to signal their intentions. In the current study, we find less flexibility for the relative timing of referential gestures and lexical items with which they share semantic material in the context of Keeps than in other locations in conversation. This is consistent with the theory that floor-holding must be carefully organized, planned, and signaled to interlocutors, whereas other turn-taking behavior may depend to a larger extent on default expectations such as those proposed by Sacks et al. (1974).

We also find that the most consistent patterns of gesture-word timing arise between onsets, and conversely, more variation between speakers is found between gesture and word offsets. It is possible that the timing of onsets is more important or in some way more salient to listeners than the offsets. Conversely, our between-speaker differences may be due to their different roles in the conversation; speaker TN is conducting a mock job interview, and TB is being interviewed, creating a clear power differential between them. A wider variety of conversational types must be investigated in order to elucidate whether differences between them are down to personal behavioral preferences or some feature of the conversational setting.

Conclusion

In a case study investigating the timing of referential hand gestures with the onset and offset of their related lexical items, we find that, just as speakers use more global gesture-speech alignment patterns to contribute to turn-taking signaling, different alignment patterns also arise on the level of the individual lexical item and the related gesture stroke. Further cross-linguistic and cross-situation research is necessary to determine to what extent these effects are consistent versus speaker- or situation-dependent.

Acknowledgements

This research was supported by the grant “Identifying Intersections between Prosody, Gesture, and Conversation” (GO 3063/1-1, PE 2879/1-1, ZE 1178/1-1), from the German Research Foundation.

References

Auer, P. (1996). On the prosody and syntax of turn-continuations. In E. Couper-Kuhlen & M. Selting (Eds.), *Prosody in conversation: interactional studies* (pp. 57.100). Cambridge, UK: Cambridge University Press.

Boersma, P., & Weenink, D. (2018). *Praat: doing phonetics by computer*. Retrieved from http://www.praat.org/.

Edlund, J., & Beskow, J. (2007). Pushy versus meek - using avatars to influence turn-taking behaviour. In *Proceedings of Interspeech 2007*, Antwerp, Belgium.

Edlund, J., & Beskow, J. (2009). Mushypeek: a framework for online investigation of audiovisual dialogue phenomena. *Language and Speech*, 52, 351-367.

Edlund, J., Beskow, J., Elenius, K., Hellmer, K., Strömbergsson, S., & House, D. (2010). Spontal: a Swedish spontaneous dialogue corpus of audio, video and motion capture. In *Proceedings of LREC 2010*, Valetta, Malta.

ELAN (Version 5.4) [Computer software]. (2019). Nijmegen: Max Planck Institute for Psycholinguistics. Retrieved from https://tla.mpi.nl/tools/tla-tools/elan/

Heldner, M., & Włodarczak, M. (2015). Pitch slope and end point as turn-taking cues in Swedish. *In Proceedings of ICPhS*, Glasgow, Scotland (pp. 10-15).

Kane, J., Yanushevskaya, I., de Looze, C., Vaughan, B., & Ní Chasaide, A. (2014). Analysing the prosodic characteristics of speech-chunks preceding silences in task-based interactions. In *Proceedings of 15th Interspeech*, Singapore (pp. 333-337).

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge, UK: Cambridge University Press.

Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., & Den, Y. (1998). An analysis of turn-taking and backchannels based on prosodic and syntactic features in Japanese Map Task dialogs. *Language and Speech*, 41, 295-321.

Local, J., Kelly, J., & Wells, W. H. G. (1986). Towards a phonology for conversation: turn-taking in Tyneside English. *Journal of Linguistics*, 22, 411-437.

McNeill, D. & Levy, E.T. (1982). Conceptual representations in language activity and gesture. In Jarvella, R.J. & Klein, W. (Eds.) *Speech, Place, and*

Action. New York: Wiley & Sons, pp. 271-295.

Mondada, L. (2007). Multimodal resources for turn-taking: pointing and the emergence of possible next speakers. *Discourse Studies*, 9(2), 194-225.

Prieto, P., Cravotta, A., Kushch, O., Rohrer, P.L., & Vilà-Giménez, I. (2018). Deconstructing beat gestures: a labelling proposal. *9th International Conference on Speech Prosody*, Poznań, Poland (201-205).

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organisation of turn-taking for conversation. *Language*, 50(4), 696-735.

Schaffer, D. (1983). The role of intonation as a cue to turn taking in conversation. *Journal of Phonetics*, 11, 243{57.

Schmidt, T. (2014). The research and teaching corpus of spoken German—FOLK. In *Proceedings of LREC 2014*, European Language Resources Association (ELRA).

Sikveland, R. O., & Ogden, R. (2012). Holding gestures across turns: moments to generate shared understanding. *Gesture*, 12(2), 166-199.

Streeck, J., & Hartge, U. (1992). Previews: Gestures at the transition place. In P. Auer & A. di Luzio (Eds.), *The Contextualization of Language* (pp. 135-158). Amsterdam: Benjamins B.V.

Zellers, M. (2017). Prosodic variation and segmental reduction and their roles in cuing turn transition in Swedish. *Language and Speech*, 60(3), 454-478.

Zellers, M., Gorisch, J., House, D., & Peters, B. (submitted). Hand gestures and pitch contours and their distribution at possible speaker change locations: a first investigation.