Prosodic manifestations of conversational roles in two types of collaborative tasks

Tatiana Kachkovskaia, Pavel Kholiavin, Anna Mamushina, Alla Menshikova

Department of Phonetics, St Petersburg University, Russia

https://doi.org/10.36505/ExLing-2021/12/0033/000506

Abstract

The acoustic features of the speaker's voice are liable to change due to a number of factors, such as success of communication, social distance between the interlocutors etc., and conversational role. The paper aims to find out how pre-defined roles of leader and follower are manifested in the speakers' prosodic features depending on the task they perform: map task vs. card-matching game. In general, information giver tends to speak louder and with higher maximum F0; temporal changes between roles are speaker-dependent. The two types of tasks differ in the prosodic manifestations of conversational roles: in map task the roles of giver and follower differ much more consistently; in card games speakers rarely reveal prosodically well-defined roles.

Keywords: speech prosody, dialogue, role-switching, map task, speaker variability.

Introduction

Much of the research of dialogue speech uses collaborative tasks to initiate active involvement of both interlocutors into the conversation. Among the most popular tasks are various types of card games, and the popular map task, where one speaker needs to describe a route drawn on a map to his/her interlocutor. It is often the case with collaborative tasks that speakers play either the role of information giver (the one who describes something, interviews the other speaker etc.), or the role of information follower (the one who listens to a description, is interviewed etc.). It has been shown that conversation roles are indeed a source of intra-speaker variability. It has been shown that information givers have longer turn duration [Forsyth et al. 2007] and more speech disfluencies [Bull, Aylett, 1998]; pause duration at turn transitions is greater when the turn is passed from the giver to the follower [Branigan et al. 1999]; between roles there are differences in speech rate [Karpiński et al. 2014]. Including the parameter of role in the analysis enables the researcher to observe curious interactions between role and gender, to trace in more detail the strategies of speech entrainment.

In this research we are estimating the influence of conversational roles on the speakers' prosodic features. As an example of a highly role-dependent collaborative task we take the classical map task. For comparison, we apply the same procedure to a card game which is presumably less role-dependent.

Material

This research is based on the corpus of Russian dialogue speech SibLing [Kachkovskaia et al. 2020]. The basic set of speakers consists of 10 pairs of same-gender siblings aged 23 to 40, and each of them was recorded in five dialogue settings: with the other sibling, a close friend of the same gender and similar age, a stranger of the same gender and similar age, a stranger of the opposite gender and similar age, and a stranger of the same gender, greater age and higher job position ("boss"). In total, the corpus contains 90 dialogues.

The card matching game was based on searching for similarities in two decks of ten whimsical pictures. The speakers took turns to describe their picture; thus, they were supposed to swap roles several times. In map task, the interlocutors were asked to guide each other through a set of schematic maps, changing the roles 4 times. The maps were not identical—some of the objects were swapped or replaced, for longer discussion. The recorded dialogues were segmented into subdialogues, each containing a discussion of a single route or card. In map tasks, speakers usually had clearly defined roles. In card games, roles could be distinguished well only at the beginning; then, the discussion became more like a free conversation and, as a result, more role-neutral. This is why for the card matching task we only analysed the first two card descriptions.

Method

For each turn in a recording, the following features were extracted: mean and maximum F_0, mean and maximum loudness, speech tempo. We removed all the turns shorter than 1.2 s from the analysis, to avoid the noise in the data that could be brought by short or mistakenly detected turns. F_0 was computed using REAPER software (https://github.com/google/REAPER); for loudness we implemented the formula described in GeMAPS [Eyben et al. 2016]. Speech tempo was computed in syllables per second based on orthographic transcriptions. Each hesitation was counted as 3 syllables.

A series of Welch's two-tailed independent t-tests was run to analyze the significance of differences in feature values between different roles. H_0 was stated as follows: there is no significant difference between the ways a person speaks as leader and as follower. For each speaker in each dialogue and for each task, t-tests were run to test this hypothesis for each of the prosodic features.

Results

Figure 1 illustrates the differences between role switching in card game and map task for maximum F_0 and speech tempo. The data are grouped with respect to speaker's gender and the interlocutor's relationship to the speaker.

Statistical analysis has shown that there are much more dialogues with significant difference between the roles within the map task than within the card game. Maximum F_0 is typically higher in leaders than in followers; the

difference is statistically significant in 77 % cases in map tasks and only in 13 % cases in card games. Mean F_0 is a weaker cue which hardly differs across roles at all. This is a reasonable finding, as speakers can easily change their maximum F_0 but not the minimum; as a result, mean F_0 has smaller variation. Speech tempo differs across roles more rarely than maximum F_0: 36 % cases in map task and 8 % cases in card games. But more importantly, some participants spoke faster as leaders (in map tasks, 26 %), while some—slower (in map tasks, 10 %). Loudness is generally higher in leaders than in followers. Maximum loudness shows significant differences across roles in 99 % cases in map tasks and 30 % cases in card games; mean loudness—in 91 % and 28 %, respectively.

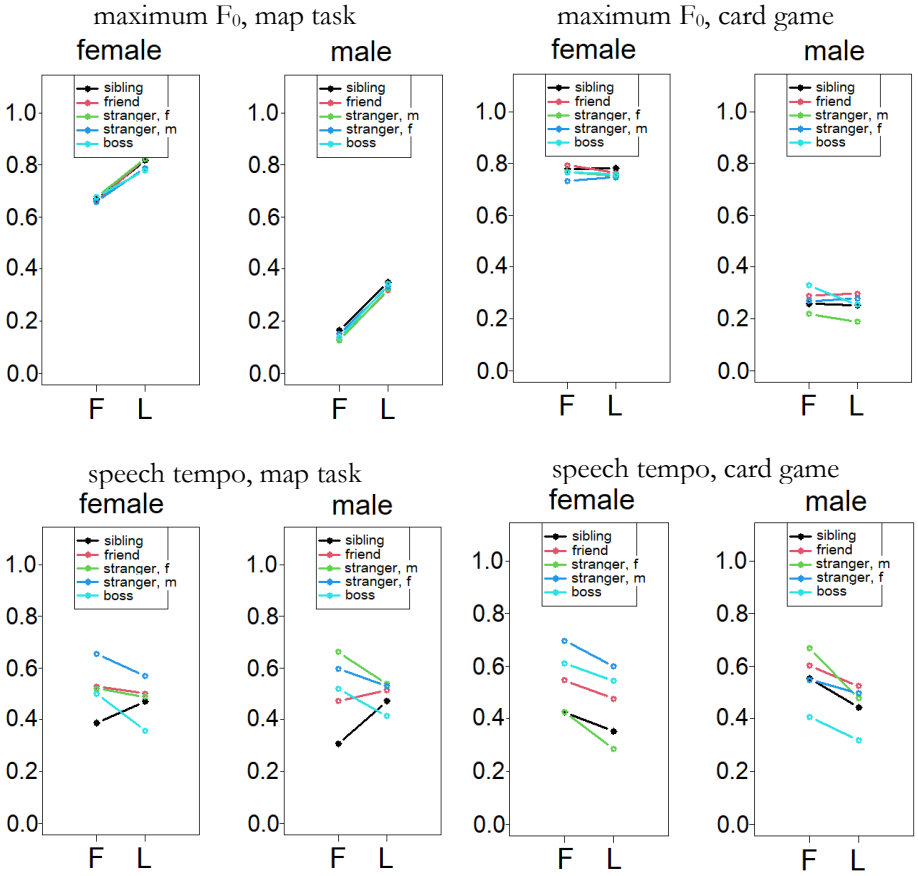

Figure 1. Relative differences in prosodic features across conversational roles (follower, F, vs. leader, L) in map tasks and card games. Various colours refer to types of relationship between the speaker and his/her interlocutor.

Discussion

When the collaborative task requires well-defined conversation roles, the interlocutors indeed speak differently according to whether they need to behave as leaders or followers. The card-matching game initiates a different type of speakers' behaviour: while in the map task the interlocutors play the "externally imposed" roles, in the card game they seem to behave more naturally. In the latter case there might still be some hierarchy, but this is probably due to the actual speaker's roles. In fact, determining the speakers' roles and hierarchy in a natural dialogue is not an easy task by itself, as relying on various types of questionnaires only provides subjective data. Knowing about the changes in acoustic features across roles could help the researchers solve this problem.

Speakers' behaviour in terms of expressing conversational roles might depend on the relationship between the interlocutors. This can be seen in Figure 1 for tempo in map task, where dialogues between siblings show opposite tendencies compared with other dialogues. A thorough analysis of all possible factors will be the next step of this research.

Acknowledgements

This work is supported by the Russian Science Foundation, grant 19-78-10046 "Phonetic manifestations of communication accommodation in dialogue".

References

Branigan, H., Lickley, R., McKelvie, D. 1999. Non-linguistic influences on rates of disfluency in spontaneous speech. Proc. 14th International Congress of Phonetic Sciences, San Francisco, USA.

Bull, M., Aylett, M. 1998. An analysis of the timing of turn-taking in a corpus of goal-oriented dialogue. Proc. 5th International Conference on Spoken Language Processing, 1175–1178, Sydney, Australia.

Eyben, F., Scherer, K. R., Schuller, B., Sundberg, J., André, E., Busso, C., Devillers, L. Y., Epps, J., Laukka, P., Narayanan, S. S., Truong, K. P. 2016. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. IEEE Transactions on Affective Computing 7 (2), 190–202.

Forsyth, R., Clarke, D., Lam, P. W. Y. 2007. Timelines, transcription and turn-taking: A corpus-driven approach to simultaneous speech. Proc. 4th Corpus Linguistics conference, University of Birmingham, UK.

Kachkovskaia, T., Chukaeva, T., Evdokimova, V., Kholiavin, P., Kocharov, D., Kriakina, N., Mamushina, A., Menshikova, A., Zimina, S. 2020. SibLing corpus of Russian dialogue speech designed for research on speech entrainment. Proc. 12th Language Resources and Evaluation Conference, 6556–6561, Marseille, France.

Karpiński, M., Klessa, K., Czoska, A. 2014. Local and global convergence in the temporal domain in Polish task-oriented dialogue. Proc. 7th International Conference on Speech Prosody, 743–747, Dublin, Ireand.