Accommodation of Backchannels in Spontaneous Speech

Antje Schweitzer and Natalie Lewandowski

Institute for Natural Language Processing, Stuttgart University, Germany
{antje.schweitzer,natalie.lewandowski}@ims.uni-stuttgart.de

We present first results from a project on phonetic convergence in spontaneous speech. Convergence is the process of accommodating one's style of speech to that of an interlocutor. Here, we examine 24 spontaneous conversations between female speakers on topics of their choice. Each dialog lasted approx. 25 minutes. Participants wore head-set microphones and could see each other through a transparent screen. There were 8 speakers, and each talked to 6 different interlocutors.

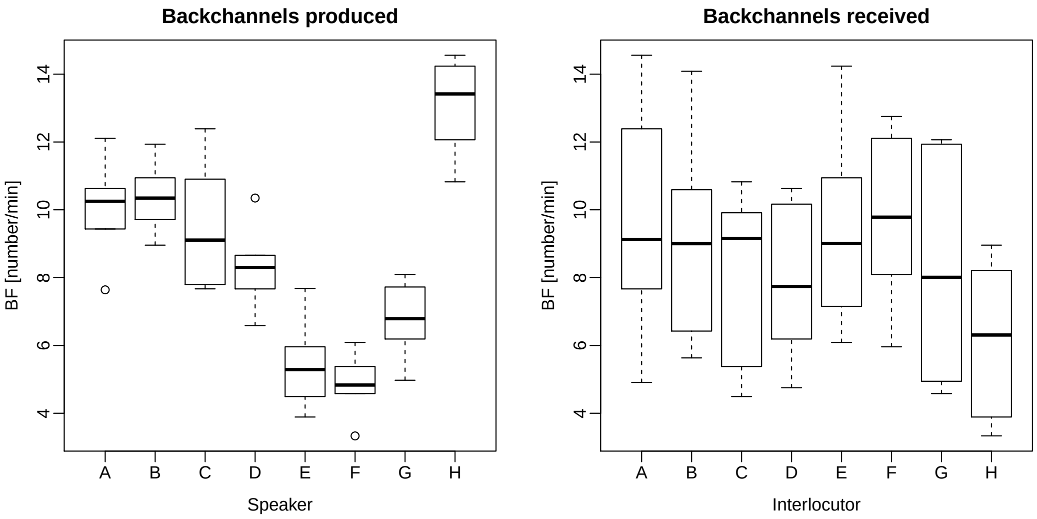
First analyses of turn-taking behavior were carried out using annotations which were generated completely automatically using Praat's (Boersma and Weenink, 2011) silence detection. We automatically identified backchannels (Yngve, 1970), i.e. short utterances produced by the listener which do not serve to interrupt the speaker's turn but serve as feedback for the speaker, such as in English "uh hum", "yeah", "o.k.", etc. We assume that each utterance of a speaker which is shorter than one second and which occurs in between utterances of the other speaker is a backchannel. We calculated backchannel frequency (BF) as the number of backchannels speakers produced in each dialog normalized by interlocutor vocalization duration. All statistical analyses presented here were conducted using R (R Development Core Team, 2011).

We were interested in speaker-specific as well as interlocutor-specific effects. Speaker-specific effects would suggest that speakers differ in their BFs. Interlocutor-specific effects, on the other hand, would indicate that speakers adjust BF depending on their interlocutor, i.e. they are accommodating (either converging to or diverging from) their interlocutor. We verified that BFs were approximately normally distributed and two Levene tests indicated no differences in variances between speakers or between interlocutors. We first ran a between-subjects analysis of variance with two factors without factor interaction (there is only one dialog for each combination of speaker and interlocutor). We found a significant speaker effect (F(7,33)=38.2, p=0.0000), indicating that BF is indeed highly speaker-specific. The interlocutor effect was also clearly significant (F(7,33)=3.9, p=0.003). This confirms that speakers differ in their BFs, and that they accommodate their BF depending on interlocutors. However, it does not indicate the direction of the effect—do they adjust their BF towards interlocutors (i.e., do they converge) or away from interlocutors (i.e., do they diverge)?

The left panel of fig. 1 shows BF by speaker, i.e., each box represents the variability in a speaker's BFs across all her six conversations. The right panel indicates BF by interlocutor, i.e. each box represents the variability in BFs that interlocutors received. It is clearly visible that BF is speaker-specific while the differences between the BFs that interlocutors received are less pronounced: the (interlocutor-specific) ranges in the right-hand graph are less well-separated than the (speaker-specific) ranges in the left-hand graph. Still, there are differences even in the right-hand graph. As for the direction of the accommodation, examine, for instance, speaker H. She produced the highest BFs across her dialogs (fig. 1, left, speaker=H). Interestingly, she received fairly low BFs from her interlocutors throughout (fig. 1, right, interlocutor=H), i.e., they diverged. Similarly, speaker F produced the lowest BFs, but received the BFs with the highest median from her interlocutors. For other speakers, BFs produced and BFs received match better.

It is well accepted that the degree of accommodation (and its direction) is related to social factors (e.g. Giles and Smith, 1979; Street, 1984; Pardo et al., 2012). To cater for such social factors in the present database, speakers rated their conversational partner (in terms of likeability, competence, etc.) after each conversation. We can assess the correlation between these mutual ratings and the BFs by fitting a linear model with BF as dependent variable and the mutual ratings as predictors. We found that the more competent or likeable speakers rated their interlocutors, the higher the BFs they produced (competence: t(46)=4.21, p=0.0001, slope=0.71, adjusted R^2=0.26; likeability: t(46)=3.64, p=0.0007, slope=0.61, adjusted R^2=0.21). Interestingly, the symmetric effect was not present: speakers who produced higher BFs were not rated as more likeable (t(46)=-0.62, p=0.54) or

Figure 1: Backchannel frequencies (BFs) by speaker (left) and by interlocutor (right).



more competent (t(46)=-1.95, p=0.058, slope=-0.37) by their interlocutors. Quite the contrary, if we count this last effect as marginally significant, it shows the inverse correlation: the slope is negative. Thus, if anything, speakers who produced higher BFs were rated as less competent by their interlocutors. This (marginally significant) second finding is in accordance with results by Jurafsky et al. (2009) and Gravano et al. (2011). The first finding, the positive correlation between how likeable and competent speakers rate their interlocutors and the backchannel frequency that they produce is a new finding, at least to our knowledge. In any case, it clearly corroborates the assumption that social factors contribute to accounting for the degree of accommodation in conversations: competence, for instance, would explain approx. 26% of the variance observed in BFs, as can be inferred from R^2 for the first regression model.

In conclusion, this first study shows that there are accommodation effects in this corpus, and that the social ratings collected do serve to capture aspects relevant for accommodation. In the future, we will look at many more parameters, especially more fine-grained ones. Also, we are interested in the dynamic aspects of convergence—we will try to assess how immediately the effects show up in conversations, and their scope.

References

Boersma, P. and Weenink, D. (2011). Praat: doing phonetics by computer (version 5.2.26) [computer program]. Retrieved from http://www.praat.org.

Giles, H. and Smith, P. M. (1979). Accommodation theory: Optimal levels of convergence. *Language and Social Psychology*, pages 45–65.

Gravano, A., Levitan, R., Willson, L., Beňuš, Š., Hirschberg, J., and Nenkova, A. (2011). Acoustic and prosodic correlates of social behavior. In *Proceedings of Interspeech 2011*, pages 97–100.

Jurafsky, D., Ranganath, R., and McFarland, D. (2009). Extracting social meaning: Identifying interactional style in spoken conversation. In *Human Language Technologies: The 2009 Annual Conference of the North American Chapter of the ACL*, pages 638–646.

Pardo, J. S., Gibbons, R., Suppes, A., and Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, 40(1):190–197.

R Development Core Team (2011). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0.

Street, R. (1984). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, 11(2):139–169.

Yngve, V. H. (1970). On getting a word in edgewise. In *Papers from the Sixth Regional Meeting of the Chicago Linguistic Society, April 16-18, 1970*, pages 567–578. Univ. of Chicago, Dept. of Linguistics, Chicago.