Alignment in Dialogue: Effects of Visual versus Verbal-feedback

Kerstin Hadelich
Department of Psycholinguistics
Saarland University
Germany
hadelich@coli.uni-sb.de

Holly P. Branigan
Department of Psychology
Edinburgh University
U.K.
holly.branigan@ed.ac.uk

Martin J. Pickering
Department of Psychology
Edinburgh University
U.K.
martin.pickering@ed.ac.uk

Matthew W. Crocker
Department of Psycholinguistics
Saarland University
Germany
crocker@coli.uni-sb.de

Abstract

It has been shown that restrictions on feedback in communicative tasks have an important impact on how speakers ground their communicative acts and the effectiveness of their communication. Generally speaking, the more interlocutors are allowed to interact, the quicker they solve communicative tasks, and the quicker they converge at a linguistic level on referring expressions for objects under discussion. Whereas the effects of verbal feedback have so far been mainly investigated with respect to linguistic measures, the effects of non-verbal feedback have been thought of as mainly influencing a more affective component or the outcome (efficiency) of communication. However, recent research has shown that visual-feedback (in terms of a shared work space) also has an effect on the smoothness and effectiveness of linguistic communication.

In our study we investigated the different effects of visual and verbal-feedback on alignment in a communicative task. In addition to commonly used measurements like the number of words of referring expressions, we also computed the lexical overlap of subsequent descriptions. We found that visual feedback also has effects on linguistic measures, and that differences in communication related to visual and verbal feedback do not necessarily show up in relatively superficial measurements such as number of words per turn.

1 Introduction

In investigating human communication, mostly task-oriented dialogues have been used. These offer the advantage that, on the one hand, participants are free to talk, but on the other hand, topic and goals of the communication are constrained by the specific task at hand. A wide variety of experiments on task-oriented dialogues have been carried out. One important issue that has been addressed is the difference that modalities used by either the speaker or the listener make on communication. In order to tackle these differences, these tasks have been conducted using different feedback conditions as variables.

One line of investigation focuses on the effects of different (verbal) task conditions on linguistic parameters. For example, the issue of coordination in the making of mutually agreeable references was addressed in a number of studies (e.g. Anderson et al., 1991; Boyle, Anderson and Newland, 1994; Clark and Wilkes-Gibbs, 1986; Clark and Krych, 2004; Horton and Keysar, 1996; Krauss and Weinheimer, 1964; Schober and Clark, 1989; ...). Clark and Wilkes-Gibbs (1986), for instance, conducted an experiment in which two participants had to arrange a set of abstract shapes (i.e., tangrams) in a linear order. One of the two participants was asked to give instructions in form of descriptions whereas the second participant was the listener who sorted the tangrams. The shapes were abstract in order to induce negotiations of names for the figure under discussion. The degree of feedback was manipulated reaching from full verbal-feedback to no-feedback. Clark and Wilkes-Gibbs measured the effects of the different feedback conditions, for example, in terms of the number of words used per referring expression.

Another line of investigation deals with the effects of visual-feedback on communication. Visual-feedback is thereby addressed either in terms of the effects of visual contact, i.e. mutual gaze or a shared visual scene, or the effect of the transmitting channel (e.g., Boyle, Anderson and Newland, 1994; Anderson, 2004; Clark and Krych, 2004; De Ruiter et al. 2003; Drolet and Morris, 2000; ...). De Ruiter et al. (2003), for example, had subjects perform a communication task in the *spatial logistics task* (SLOT), a psycholinguistic version of the so-called *social dilemma scenario*. In SLOT two participants have to negotiate a route through a map that meets certain optimisation criteria. In their experiment, the visual information of the scene was shared across all conditions. De Ruiter et al. looked at the effects of the presence and absence of eye contact on the outcome of the task. In one condition they used a one-way mirror that only allowed asymmetric visual contact. De Ruiter et al. found that in this condition negotiation times increased significantly but the successful outcome of the task was not affected. This result is consistent with findings in earlier work (e.g., Drolet and Morris, 2000). Remarkably in this respect, Anderson (2004) reports that in one map task experiment (Anderson et al., 1991) only 30 % of words were actually uttered in the time span of mutual gaze.

Most of these problem-solving tasks are asymmetric by virtue of the way in which roles are assigned to (the two) interlocutors, such as instruction giver versus instruction receiver. Even though this design reveals obvious disadvantages when it comes to the generalisation of results, it nonetheless appears to be an approach that satisfies many of the relevant constraints.

Taken together, the evidence suggests that visual-feedback affects the way in which participants solve a task in dialogue. But apart from more general measures like efficiency and affective components (e.g. rapport), it remains unclear what influence different feedback modalities have on the linguistic dimensions of dialogue. In our study, we compared the effects of visual-feedback (shared visual information about the scene but no eye contact) versus verbal-feedback on linguistic measures of communicative success. In measuring the linguistic effects, we use the concept of alignment (Pickering and Garrod, in press) and analyse, for example, the lexical overlap in subsequent utterances.

2 Experiment

We tested 32 Edinburgh University students who received £5 each for taking part in the (30 - 60 minute) experiment. Participants were paired randomly and randomly assigned to one of the four experimental conditions.

2.1 General set-up

Participants were separated by a head-high divider. Each participant was seated in front of a monitor and given a separate mouse. Their task was to move a set of tangrams from an initial set of positions into their final positions as indicated on a given target configuration. The two boards were identical and showed all eight tangrams. However, each participant had their own individual target card with four of the eight tangrams displayed on it. Both the board to play on and the target card were displayed on the monitor (see Figure 1). Participants were asked to take turns instructing each

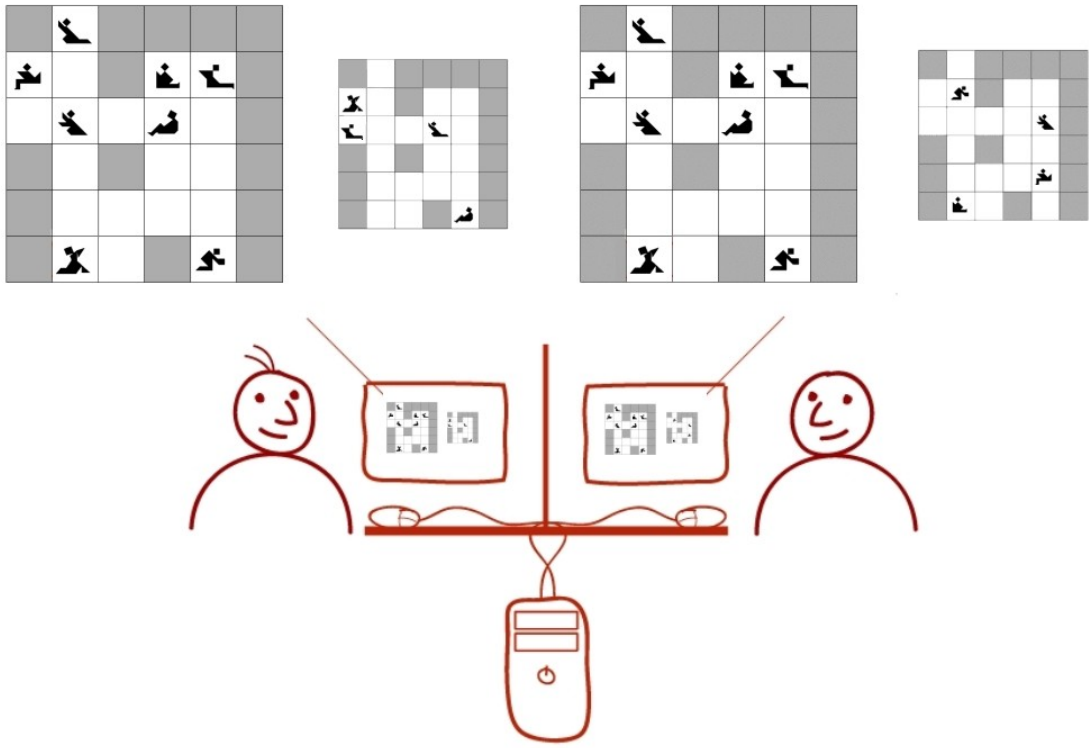

Figure 1: Schematic illustration of the experimental set-up. The board with all eight tangrams and, next to it, the target card showing the final positions of four of the tangrams as the two participants saw them on their screens.

other. This means that they alternately selected a tangram on their target card and gave instructions to the instruction receiver until this particular tangram had reached its final position. In doing so, other tangrams had potentially to be moved out of the way first. After the selected tangram had reached the target position, participants swapped roles. This sequence was repeated until all eight tangrams had reached their final positions and the target configuration was accomplished. Our aim was to approach the symmetric character of natural conversation by introducing a more dynamic role assignment. Also, the turns taken in giving instructions can be seen as an equivalent to subgoals in conversation.

Prior to running the experiment, participants completed a practice session illustrating the rules and technical features of the set-up. In this practice session we used geometric shapes instead of tangrams, to avoid giving the participants practice in the specific task.

2.2 Conditions

We varied the type of feedback that participants could give in a between-participants design. Each pair of subjects was randomly assigned to one of four conditions: *full-feedback*, *verbal-feedback*, *visual-feedback*, and *no-feedback*.

In the full-feedback condition, we allowed participants to talk freely; additionally the two monitors were connected, so that the instruction giver also could see on their screen which of the items the instruction receiver was moving, and to which position. In the only-verbal-feedback condition, participants were also allowed to talk freely. But this time their monitors were not connected, so they did not get any information about which item their partner was moving. In the only-visual-feedback condition, the instruction receiver could not give any verbal feedback, but participants

could again see what their partner was doing on their screen. Finally, in the no-feedback condition, the instruction receiver could not give any verbal feedback and the participants' monitors were also not connected.

2.3 Hypotheses

Generally, alignment takes place most effectively by the use of same channels in interaction and shows up in same representations used by interlocutors (Pickering and Garrod, in press). Participants are thus expected to prime each other in the use of, e.g., lexical items. This effect should be stronger when interlocutors are allowed to interact verbally as opposed to a more passive participation in the communication when listening to instructions. We thus expected a greater reduction of number of words and greater lexical overlap in subsequent descriptions in verbal-feedback conditions. This advantage should result in fewer disfluencies in the verbal-feedback conditions.

2.4 Analysis

We identified the first phrase of each referring expression that was delimited by intonational phrase boundaries. We analysed the number of words in a phrase in order to measure the process of convergence and additionally looked at the number of disfluencies (e.g., filled pauses, such as *uh* and *uhm*). In the conditions without verbal-feedback (i.e., visual-only and no-feedback) the descriptions could not be interrupted by the listener and thus tended to be much longer than the verbally more interactive conditions. In cutting down the descriptions into smaller, phrasal units, we increased comparability of the utterances across conditions. We also computed lexical overlap of subsequent descriptions. The *relative lexical overlap* for a description *k* was calculated by relating the number of lemmas in description *k* shared with description *k-1* to the total number of lemmas in descriptions *k* + *k-1*. As in the first description of an item in conversation the preceding description is missing, we only included descriptions two, three, and four in the analyses to compute lexical overlap.

2.5 Results

We conducted univariate ANOVAs and paired comparisons (Scheffé Posthoc Test) with participants as random factors and disfluencies, number of words used in the first phrase, and lexical overlap as dependent measures. Overall, the visual-feedback conditions differed from the verbal-feedback conditions with respect to disfluencies and relative lexical overlap, but not with respect to length of the description.

The results showed a significant main effect of condition on number of disfluencies in the referring expressions ($F(3, 1509) = 73.022$; $p < .005$). The paired comparisons revealed an effect of feedback modality: Conditions without verbal feedback (visual and no-feedback) showed significantly more disfluencies than the two conditions with verbal-feedback (full and verbal-feedback). Additionally, the absolute number of words used for the first phrase in a description also showed significant effects of condition type ($F(3, 650) = 44.996$; $p = .005$). Here, the posthoc tests revealed that in the no-feedback condition significantly fewer words per phrase were used than the other three feedback conditions (see Figure 2). The third dependent measure, relative lexical overlap, also showed significant effects of condition type ($F(3, 647) = 14.388$; $p < .05$). As in the analysis of disfluencies, there was again a significant effect of feedback modality. But this time the effect was inverted: In the two verbal-feedback conditions, referring expressions shared significantly fewer lemmas with their preceding utterance than the two conditions without verbal feedback.

3 Conclusion

The data provide only partial support for the hypothesis that verbal-feedback is more effective for alignment than visual-feedback. With respect to fluency, verbal feedback turned out to have the expected effects, i.e. in conditions with verbal feedback, utterances were more fluent than those in conditions without verbal feedback. The second commonly used measurement in the analyses of dialogue, length of referring expressions, did not reveal differences of verbal versus visual feedback. Only in the no-feedback condition, in

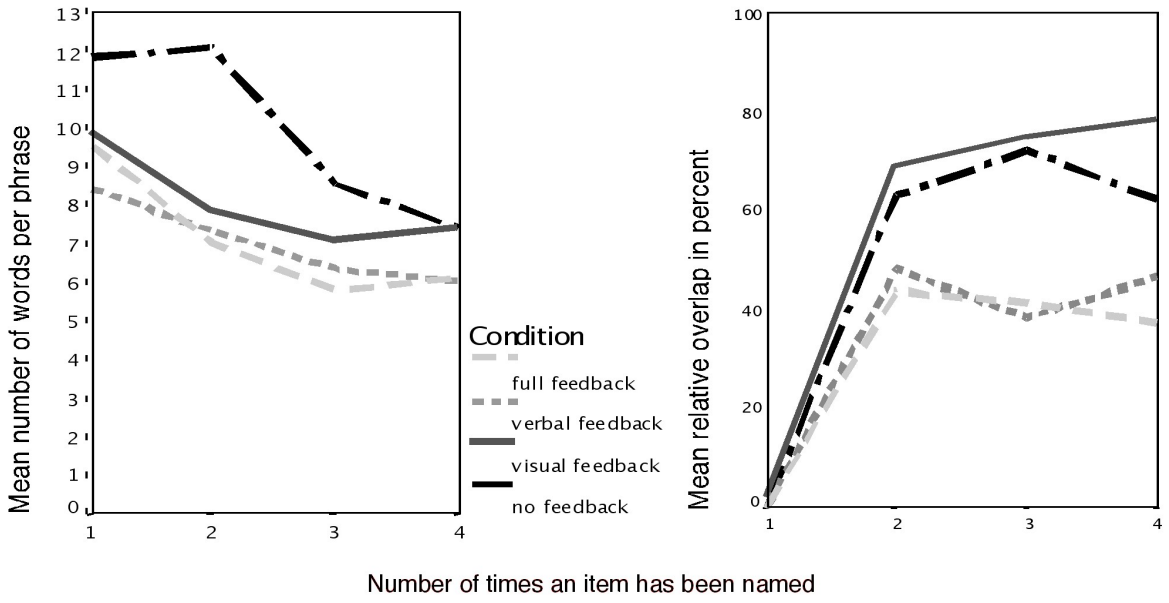

Figure 2: Mean number of words per phrase and mean relative overlap per phrase relative to the number of times an item has been named.

which participants could neither see what their partner was doing nor negotiate names for an item to be moved, did participants produce significantly longer referring expressions than in the three other feedback conditions. Noteworthy at this point is that, obviously, in task-oriented dialogues as the one described above, the type of feedback does not seem to make a difference with respect to the length of the first phrase. The more important factor appears to be the actual possibility of having feedback in communication, be it visual or verbal. Moreover, the fact that the conditions with verbal-only and visual-only feedback are not significantly different from the full feedback condition suggests that the communicative benefit on the first phrase is not larger with an increase of feedback. However, the predicted difference between visual and verbal-feedback did show up in the measure of relative lexical overlap that we computed for subsequent descriptions. Here we found that there was less overlap in the two verbal-feedback conditions than in the visual or the no-feedback condition. To some extent, this is surprising as the assumptions drawn on the basis of the alignment model pointed into the opposite direction. One way to interpret these results is to consider the overlap showing up in the verbal-feedback conditions as the automatic portion of overlap and the additional overlap in the visual-feedback conditions as stemming from other origins, such as pragmatic or situational influences or an aspect of audience design. In conditions without verbal feedback, participants have to make sure that their descriptions are understandable. This is even more the case in the no-feedback condition, because misunderstandings are much more difficult to resolve. In order to make sure that referring expressions are understandable, an appropriate strategy can be to reuse successful lemmas, which leads to a relatively big overlap. The interactive character of the verbal feedback conditions, however, offered instruction receivers the possibility to actively take part in the process of finding a name for items under discussion. Thus, subsequent descriptions in the verbal feedback conditions don't necessarily have to be driven by an automatic tendency to

align on a name, but could also show effects of this collaboration.

Taken together, we have shown that visual feedback obviously has effects not only on more general and affective components of communication but also on linguistic measures such as the number of words used in a referring expression and lexical overlap. This further highlights the fact that differences between visual and verbal-feedback are not revealed in relatively superficial measures such as the number of words and require more fine-grained measures such as degree of lexical overlap.

Acknowledgements

We would like to thank Alissa Melinger and Andrea Weber for helpful comments on the experimental design and an earlier draft of this paper. The presentational software for the study described was programmed by Alan Marshall (Edinburgh University).

References

A.H. Anderson, M. Bader, E.G. Bard, E. Boyle, G. Doherty, S. Garrod, S. Isard, J. Kowtko, J. McAllister, J. Miller, C. Sotillo, and H.S. Thompson. 1991. The HCRC Map Task Corpus *Language & Speech*, 34:351–366.

A.H. Anderson. 2004. Feedback in problem solving dialogue. Talk given on the Multi-Modal Interaction Projects Feedback in Interaction Workshop, MPI Nijmegen.

E.A. Boyle, A.H. Anderson, and A. Newlands. 1994. The effects of visibility on dialogue and performance in a cooperative problem solving task. *Language & Speech*, 37(1):1–10.

H.H. Clark and D. Wilkes-Gibbs. 1986. Referring as a collaborative process. *Cognition*, 22:1–39.

H.H. Clark and M.A. Krych. 2004. Speaking while monitoring addressees for understanding. *Journal of Memory and Language*, 50(1):62–81.

J.P. De Ruiter, S. Rossigno, L. Vuurpijl, D.W. Cunningham, and W.J.M. Levelt. 2003. SLOT: A research platform for investigating multimodal communication. *Behavior Research Methods, Instruments, and Computers*, 35(3):408–419.

M.J. Pickering and S. Garrod. in press. Toward a mechanistic Psychology of Dialogue. *Behavioral and Brain Sciences*

G. Doherty-Sneedon, A.H. Anderson, C. O'Malley, S. Langton, S. Garrod, and V. Bruce. 1997. Face-to-Face and video mediated communication: A comparison of dialog structure and task performance. *Journal of Experimental Psychology*, 3, 2:105–123.

A.L. Drolet and M.W. Morris. 2000. Rapport in conflict resolution: Accounting for how face-to-face contact fosters mutual cooperation in mixed motive conflicts. *Journal of Experimental Social Psychology*, 36:26–50.

W.S. Horton and B. Keysar. 1996. When do speakers take into account common ground? *Cognition*, 59(1):91–117.

R.M. Krauss and S. Weinheimer. 1964. Changes in reference phrases as a function of frequency of usage in social interaction: A preliminary study. *Psychonomic Science*, 1:113–114.

M.F. Schober and H.H. Clark. 1989. Understanding by addressees and overhearers. *Cognitive Psychology*, 21:211–232.