How listeners respond to speaker's troubles

Patrick G. T. Healey, Mary Lavelle, Christine Howes,
Stuart Battersby, Rose McCabe
ph@eecs.qmul.ac.uk
Queen Mary University of London,
Cognitive Science Research Group,
School of Electronic Engineering and Computer Science,
London E1 4NS, UK

Abstract

Listeners normally provide speakers with simultaneous feedback such as nods, "yeah"s and "mhm"s. These 'backchannels' are important in helping speakers to talk effectively. Two factors are known to influence when a backchannel is produced; if the speaker is looking at the listener or if the speaker is presenting new information. We investigate a third factor: whether the speaker is having trouble speaking i.e. self-repair. If dialogue is an active collaborative process then listener's responses should be especially critical when trouble is encountered. Using data from a corpus of three person dialogues we show that speaker's rate of self-repair is a better predictor of listener responses than speech rate. We also show that listeners respond strongly to speaker troubles independently of whether the speaker is looking at them. We argue that it is the points at which conversation threatens to go off-course that are most significant for coordination. **Keywords:** Gesture; repair; dialogue

Introduction

Listening in conversation is not a passive activity. As Goffman (1955) noted, what listeners do while being addressed has important consequences for the way that speakers produce their turns. Goffman distinguished between two general kinds of listener feedback; displays of attention and understanding of what is said and the signalling of interactional functions such as a desire to speak next. Yngve (1970) introduced the term 'backchannel' to describe these uses of simultanous feedback that provide speakers with concurrent information about how their turn is being received.

In a series of experiements examining the effects of listener response behaviours Bavelas and colleagues were able to show that the fluency and effectiveness of a speaker's turns depends directly on the level of feedback they are getting from their addressees (J. B. Bavelas et al., 2000; J. Bavelas et al., 2006). People telling stories to listeners who are engaged in a distractor task speak less fluently and are less compelling than those whose listeners are attending more carefully.

Given the importance of listener responses for successful interaction a key question is what prompts a listener to produce them? Many of the most common backchannel signals, such as nods and smiles, use the visual channel which avoids potential competition with concurrent speech. One common finding in the literature is that addressee responses are reliably correlated with speaker's gaze. Goodwin (1979) observed that speakers will periodically check whether addressees are attending by looking at them and if they get no response may restart or switch to a new addressee mid-turn. J. B. Bavelas et al. (2002) found that listener responses in their 'close call' story telling task were significantly more likely to occur in a 'gaze window' i.e. when a speaker is looking at a listener than when they are not.

A second common observation in the literature is that backchannels are also associated with the introduction of new information into a dialogue such as the introduction of a new referent or proposal that may warrant some signal of interim acknowledgement or acceptance before the speaker's turn is completed (J. Bavelas et al., 2006; Clark & Wilkes-Gibbs, 1986; Yngve, 1970). In this case it is the information update that prompts the use of a backchannel to signal understanding 'so far' (Goodwin, 1981).

In this paper we explore the effects of a third factor on listener responses: the degree of difficulty a speaker has in producing their turn. Few conversational turns are produced without some form of online revision or reformulation during their production. Sometimes referred to as disfluencies these self-repairs are indicative of some sort of trouble producing a turn. If conversation is a collaborative process in which each turn is co-produced (Goodwin, 1979; Clark, 1996) then this leads to the hypothesis that the points at which the speaker shows signs of getting into trouble ought to be especially critical for collaborative reponses. This paper tests this hypothesis by investigating the relationship between nodding, speech rate and repair rate in a corpus of three person dialogues.

Methods

Experimental work on listener backchannel responses has focussed only on dyadic, i.e. two person, interactions. However, natural interactions frequently involve more than two people (Goffman, 1981; Eshghi, 2009). For current purposes three-way interactions also have the practical advantage that they make it possible to compare two kinds of listener depending on who the speaker is looking at while they talk. Given the importance of speaker gaze to the production of backchannels this provides a useful opportunity to compare the responsiveness

of two fully ratified, active participants who differ only in whether they are being looked at while the speaker produces their turn. Note that this differs from the work of Schober & Clark (1989) who investigated the behaviour of side participants and overhearers whose ability to provide concurrent feedback was restricted.

Participants

Fifty four participants (30 Male, 24 Female) were recruited to the study through advertising on local community websites. Of those who responded to the advertisement, 40% participated. Participants within each group had not met prior to the study.

Procedure

Participants were brought into the laboratory in threes and seated in a triangular formation so that each participant had good visual access to each of the others (see Figure 1). The researcher read aloud a fictional moral dilemma scenario called 'the balloon task' to the seated group. The scenario states that there are four people in a hot air balloon, which is losing height and about to crash into some mountains killing all on board. One person must jump from the ballon to their certain death in order to save the other three. Participants were instructed to debate the reasons for and against each person being saved, and reach mutual agreement about who should jump. The group was provided with an opportunity to ask questions before the researcher left the interaction space and the task began. Interactions ended when participants reached a joint decision. Groups that failed to reach an agreed decision had their interaction terminated at approximately 450 seconds (7 minutes 30 seconds).

Figure 1: 2-dimensional image of participants engaged in triadic interaction, wearing the reflective markers

All interactions were recorded in a human interaction laboratory fitted with an optical based Vicon motion-capture system, consisting of 12 infrared cameras and Vicon iQ software. Participants wore a top and a cap with 27 reflective markers attached. Cameras detected the markers at 60 frames per second, resulting in a highly accurate 3D representation of participants' movements over time (see Figures 1 and 2).

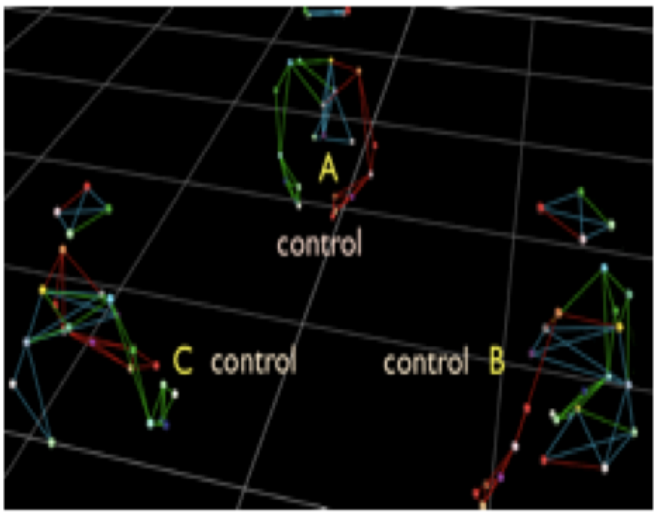

Figure 2: The wire frame representation of the interaction in 3-dimensional space

Data Analysis

For each interaction, speech was transcribed from the 2D video in the annotation tool ELAN (Crasborn & Sloetjes, 2008). These transcripts, together with the motion capture data were used to produce three measures, speech rate, rate of self repair and rate of nodding for each participant.

Measures of Self-Repair Automatic processing of the transcripts identified, for each turn, the number of words, the number of filled pauses (e.g. *er*, *um*) and the number of unfilled pauses, defined as pauses between segments of speech by the same speaker of greater than 200 milliseconds (following e.g. Zellner, 1994, a.o.). Since self-repairs often involve the repetition of words, usually close together, a normalised within-turn repeated words value was calculated, by identifying repeated words in a turn and the distance between them and applying a decay function. Examples of turns including self-repetition, and their word repeat value are shown below, from a low repeat score in example 1 to a high repeat score in example 3. Repeated words are shown in bold, and their repetition in italics.

(1) sort of **long** so they're usually about that *long* I think [0.17]

(2) Trust me **his wife if he's** *if he's* a pilot *his wife* knows how to do it [1.25]

(3) **And** *and* **they** *they* **said that** she *said that they* emptied the balloon to make it lighter [2.98]

To check validity, this measure was also calculated on a corpus of 52 clinical dialogues which had been hand-annotated for self-repair (McCabe et al., in preparation). For the 15,191 turns analysed, the within-turn repeated words measure was positively correlated with the hand-annotated self-repair measure ($r = 0.57, p < 0.001$) and is therefore used as an index of self-repair. All values were normalised by number of frames in the turn, and mapped to the frame-by-frame motion capture data.

Nodding Head movement was derived from the vertical movement of participants front left head marker. Head nodding was approximated in a two-step process. Firstly, low frequency movements (1Hz and below) and high frequency movements (4Hz and above) were eliminated, in accordance with those described as the parameters of normal head movement in the British Journal of Ophthalmology (Gresty et al., 1976) and fall within the range of ordinary head movement as described by Hadar et al. (1983). Secondly, in line with previous studies (Cerrato & Svanfeldt, 2006), head nods were identified as vertical movements at a speed >0.3 mm/frame, with 7 frames between the top and bottom of the movement.

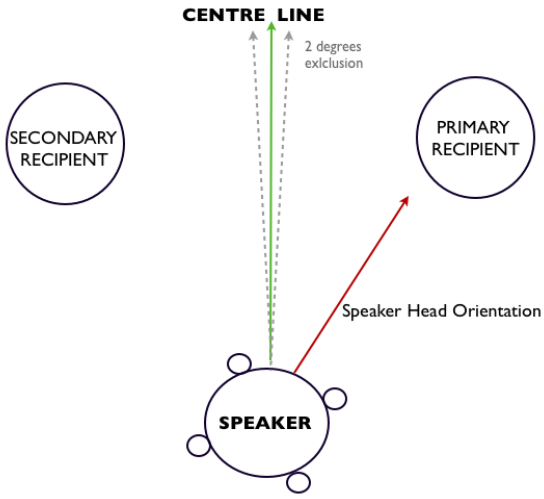

Figure 3: Indexing dialogue role through speaker head orientation

Speaker Orientation and Recipient role The speech transcript was synchronized with the 3D motion capture data, identifying the identity of the speaker(s) in each frame of interaction. In order to identify the speaker's primary addressee at each point in the dialogue the technique described in Healey & Battersby (2009) was used. For each frame of data the speakers' head orientation is calculated using the coordinates of their four head markers. The orientation of the speakers' head is compared to a centre line falling between the speakers' two interacting partners, bisecting the interaction space (Figure 3). Head orientations falling within two degrees of the centre line are excluded. If the speaker's head orientation falls on one side of this line the person on that side is coded as the primary recipient i.e. the person the speaker is primarily orienting to at that point in the dialogue. The other participant is coded, by default, as the secondary recipient. The identity of the speaker (based on hand annotated speech) and the primary and secondary recipients (based on speaker head orientation) is coded for each frame of data. Although in principle head orientation is independent of gaze direction it is nonetheless a reliable indicator of speaker's attention and gaze, especially in multi-party dialogue (Healey & Battersby, 2009; Jokinen et al., 2010; Loomis et al., 2008).

Results

Following Boker et al. (2002), windowed cross-correlations were used to determine the degree of coordination between the head nodding of each participant (i.e. speaker, primary recipient and secondary recipient) and the speaker's speech and repair rates. This method directly compares the rates of speakers' speech and repair at each frame with the head movement of each participant on a lagged frame-by-frame basis within each 30-second window providing: (i) the correlation between speakers' rate of self-repair/speech and participants' nodding, and (ii) the temporal offset at which they occur. Consecutive windows were overlapped to minimize the chance of significant correlations being undetected. Windowed cross-correlation analyses assume local stationarity within each window. Although this may not always be the case, any violations will produce a downward bias of correlation and lag, providing a conservative measure of the magnitude of the effects (as discussed in Boker et al., 2002).

Figure 4 shows the results of the cross-correlation of nodding with speech rate, at lags of up to ±240 frames (4 seconds). At zero offset speakers nod most, primary recipients nod less and side participants nod least. The Friedman comparisons in Table 1 shows this global pattern of differences between roles is reliable. As the Figure shows, the difference in roles is greatest at zero offset. This is consistent with a pattern in which speakers nod most, primary participants produce some feedback through nods and secondary participants supress their nodding, as indicated by the negative correlation. Since all participants take all roles in this task these effects are only due to differences in who the speaker is looking at.

The cross-correlation of nodding with repair rate, illustrated in Figure 5, shows a different pattern of timing

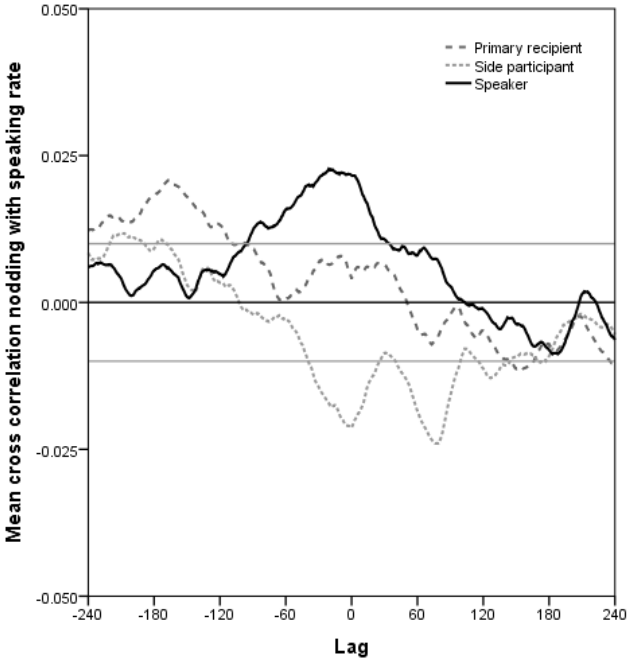

Figure 4: Cross-Correlation of Speech Rate and Rate of Nodding. Horizontal grey lines indicate the 95% confidence interval.

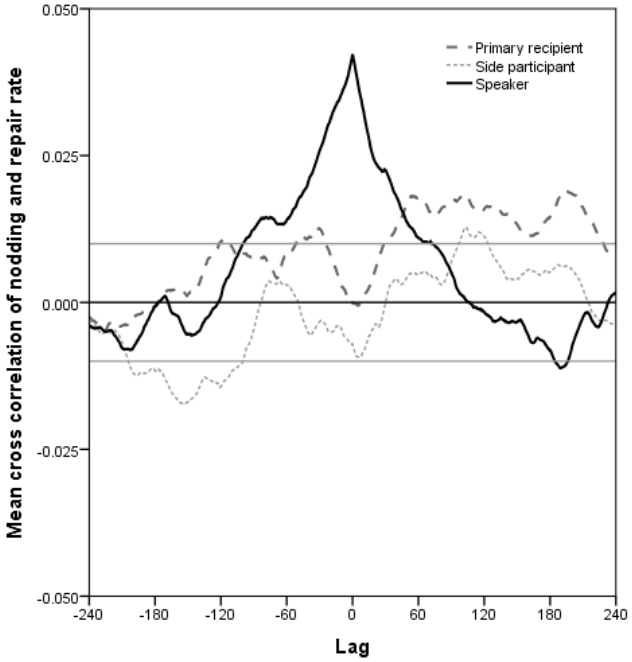

Figure 5: Cross-Correlation of Repair Rate and Rate of Nodding. Horizontal grey lines indicate the 95% confidence interval.

and level of responses to repair rate than to speech rate. As Table 1 shows speakers still nod more than primary or side participants in turns that include repairs, however both people in the recipient role at the time the of the repair nod significantly more than they would otherwise. Especially in the 1-3 second offset, i.e. towards the end of the turn involving a repair.

Pairwise comparison			Friedman's test stat		
			Raw	Std	p
Repair Rate	Speaker	Primary	-0.784	-6.497	< 0.001
	Speaker	Side	1.042	8.634	< 0.001
	Primary	Side	1.825	15.131	< 0.001
Speech Rate	Speaker	Primary	0.574	4.757	< 0.001
	Speaker	Side	1.757	14.562	< 0.001
	Primary	Side	1.183	9.806	< 0.001
Repair vs Speech	Speaker		-0.287	-2.378	0.261
	Primary		1.071	8.875	< 0.001
	Side		0.428	3.550	0.006

Table 1: Non-parametric test results for cross-correlations by role and speech or repair rate pairwise comparisons

Friedman pairwise comparisons show no reliable difference in speakers nodding as predicted by speech rate or repair rate but both recipient roles show a significantly stronger response to repair rate. Secondary participants, in particular, shift from suppressing their nodding behaviour while the speaker is addressing someone else to a profile much more similar to that of a primary participant especially at offsets of between 1 and 3 seconds.

Discussion

Despite the fact that all three people involved in the balloon task dialogues are active, ratified participants who are free to respond at any time, the results indicate that there are clear differences in levels of responsiveness depending on who the current speaker is attending to as indexed by theier head orientation. This is consistent with previous work by Goodwin (1979) and J. B. Bavelas et al. (2000); J. Bavelas et al. (2006) who emphasise the importance of speaker gaze in eliciting listener responses.

The results reported here extend existing findings in two ways. Previous experimental work has focussed on the behaviour of listeners in dyadic i.e. two-person dialogues. Here we extend this to three person dialogues. A pragmatic feature of three-person dialgoues is that it becomes harder to judge who is speaking to whom and, as a result, more difficult to co-ordinate the roles of speaker and addressee.

Our results demonstrate that in this context there are concurrent differences in people's levels of responsiveness that depend on whether the speaker is currently oriented to them or to someone else in the conversation; indepen-

dently of what is being said. This replicates findings reported by Healey & Battersby (2009), for a different corpus, which indicated that listerners who are not oriented to by the speaker, i.e. secondary recipients, are normally less responsive than primary recipients.

Consequently, it is not merely exposure to the content of what is said that determines responsiveness. Interestingly, these results also show for the first time that secondary participants actually suppress non-verbal feedback i.e. their head movements are substantially negatively correlated with the speaker's speech. It appears likely that this is because they are, in a sense, actively displaying their non-recipiency.

Importantly, the present results also suggest the influence of a new factor on overt levels of response; self-repair or speaker troubles. Although there is no overall effect on Speaker's nodding, Listeners in both the primary and secondary recipient roles respond more strongly to turns in which there is evidence that the speaker is having trouble formulating or articulating their message. This is significant, in part, because self-repairs are relatively common, occuring in at least a third of turns in natural dialogue even on conservative estimates (Colman & Healey, 2011). The effect is more marked for secondary recipients who switch from suppressing their responses to producing a profile much closer to that of the primary recipient.

The implication of these differences in patterns of repsonsiveness is that it listener feedback is primarily organised around the successful construction of a turn, not the content of that turn. This strengthens the view that conversation is an active, collaborative process in which people make concerted use of the resources available to them, including speech, gesture and head movements, to produce each turn. However, it also suggests that these resouces are most actively used to help speakers recover from problems in the production of their turn and not, as normally assumed, for acknowledging or 'grounding' new information.

Acknowledgments

We gratefully acknowledge the support of the Engineering and Physical Sciences Research Council (EPSRC) Doctoral Training Programme (EP/P502683/1).

References

Bavelas, J., Coates, L., & Johnson, T. (2006). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, *52*(3), 566–580.

Bavelas, J. B., Coates, L., & Johnson, T. (2002). Listener responses as a collaborative process: The role of gaze. *Journal of Communication*, *52*(3), 566–580.

Bavelas, J. B., Coates, L., Johnson, T., et al. (2000). Listeners as co-narrators. *Journal of personality and social psychology*, *79*(6), 941–952.

Boker, S., Rotondo, J., Xu, M., & King, K. (2002). Windowed cross-correlation and peak picking for the analysis of variability in the association between behavioral time series. *Psychological Methods*, *7*(3), 338.

Cerrato, L., & Svanfeldt, G. (2006). A method for the detection of communicative head nods in expressive speech. *Gothenburg papers in theoretical linguistics*(92), 153–165.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

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

Colman, M., & Healey, P. G. T. (2011). The distribution of repair in dialogue. In *Proceedings of the 33rd annual meeting of the cognitive science society* (pp. 1563–1568). Boston, MA.

Crasborn, O., & Sloetjes, H. (2008). Enhanced ELAN functionality for sign language corpora. In *3rd workshop on the representation and processing of sign languages: Construction and exploitation of sign language corpora.*

Eshghi, A. (2009). *Uncommon Ground: The Distribution of Dialogue Contexts*. Unpublished doctoral dissertation, Queen Mary University of London.

Goffman, E. (1955). On face-work: an analysis of ritual elements in social interaction. *Psychiatry: Journal for the Study of Interpersonal Processes*.

Goffman, E. (1981). *Forms of talk*. University of Pennsylvania Press.

Goodwin, C. (1979). The interactive construction of a sentence in natural conversation. In G. Psathas (Ed.), *Everyday language: Studies in ethnomethodology* (pp. 97–121). New York: Irvington Publishers.

Goodwin, C. (1981). *Conversational organization: Interaction between speakers and hearers*. New York: Academic Press.

Gresty, M., Leech, J., Sanders, M., & Eggars, H. (1976). A study of head and eye movement in spasmus nutans. *British Journal of Ophthalmology*, *60*(9), 652–654.

Hadar, U., Steiner, T., Grant, E., & Rose, F. C. (1983). Kinematics of head movements accompanying speech during conversation. *Human Movement Science*, *2*(1), 35–46.

Healey, P. G. T., & Battersby, S. A. (2009). The interactional geometry of a three-way conversation. In N. Taatgen & H. van Rijn (Eds.), *Proceedings of the thirty-first annual conference of the cognitive science society.*

Jokinen, K., Nishida, M., & Yamamoto, S. (2010). On eye-gaze and turn-taking. In *Proceedings of the 2010 workshop on eye gaze in intelligent human machine interaction* (pp. 118–123).

Loomis, J., Kelly, J., Pusch, M., Bailenson, J., & Beall, A. (2008). Psychophysics of perceiving eye-gaze and head direction with peripheral vision: Implications for the dynamics of eye-gaze behavior. *Perception*, *37*(9), 1443–1457.

McCabe, R., Lavelle, M., Bremner, S., Dodwell, D., Healey, P. G. T., Laugharne, R., ... Snell, A. (in preparation). *Shared understanding in psychiatrist-patient communication: Association with treatment adherence in schizophrenia.*

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

Yngve, V. H. (1970, April). On getting a word in edgewise. In *Papers from the 6th regional meeting of the chicago linguistic society* (p. 567-578).

Zellner, B. (1994). Pauses and the temporal structure of speech. In E. Keller (Ed.), *Fundamentals of speech synthesis and speech recognition* (p. 41-62). Chichester: John Wiley.