Proceedings of DiSS'05, Disfluency in Spontaneous Speech Workshop.
10–12 September 2005, Aix-en-Provence, France, pp. 133–138.

Disfluency & Behaviour in Dialogue: Evidence from Eye-Gaze

Hannele Nicholson[1], *Ellen Gurman Bard*[1], *Robin Lickley*[2],
Anne H. Anderson[3], *Catriona Havard*[3] & *Yiya Chen*[1]

[1] University of Edinburgh, Edinburgh, Scotland
[2] Queen Margaret University College, Edinburgh, Scotland
[3] University of Glasgow, Glasgow, Scotland

Abstract

Previous research on disfluency types has focused on their distinct cognitive causes, prosodic patterns, or effects on the listener [9, 12, 17, 21]. This paper seeks to add to this taxonomy by providing a psycholinguistic account of the dialogue and gaze behaviour speakers engage in when they make certain types of disfluency. Dialogues came from a version of the Map Task, [2, 4], in which 36 normal adult speakers each participated in six dialogues across which feedback modality and time-pressure were counter-balanced. In this paper, we ask whether disfluency, both generally and type-specifically, was associated with speaker attention to the listener. We show that certain disfluency types can be linked to particular dialogue goals, depending on whether the speaker had attended to listener feedback. The results shed light on the general cognitive causes of disfluency and suggest that it will be possible to predict the types of disfluency which will accompany particular behaviours.

1. Introduction

Types of disfluency distinguished by their form are also distinguishable by other characteristics. Repetition disfluencies are the most common in spontaneous speech [21]. In a pioneering paper, Maclay & Osgood showed that repetitions precede content words more often than function words [22]. Repetitions have been linked to strategic signalling commitment to both listener and utterance [10, 12]. The prosodic cues for repetitions are linked to certain strategies in dialogue [25]. Savova showed, however, that the prosodic cues to repetitions differ from the cues to a substitution, providing support for the notion that disfluency types have distinct sources in the cognitive processes underlying the production of speech in dialogue [26].

It is already clear that disfluencies of different types cause different processing problems for the listener. While repetitions cause less disruption than false starts [a kind of deletion disfluency] for a word recognition task, [13], repetitions are more difficult for trained transcribers to detect than false starts of the same length [20].

Disfluency has been linked to cognitive causes by Levelt [17], who proposes that some disfluencies occur for covert cognitive reasons while other disfluencies are overt corrections. Lickley found that disfluency types vary systematically across turn types whereby turns that involve planning typically involve more self-corrections than utterances which are responses to queries [18]. Replies to queries, on the other hand, tend to involve more filled pauses (ums, uhs) and repetitions in order to buy time [18]. Thus, it seems that certain types of disfluencies have already been linked to certain dialogue behaviours.

More recently, psycholinguistic studies of a speaker's eye-gaze at a visual array have revealed that speakers look at objects involved in the process of speech perception and production. [15, 28]. Speakers who made a speech error when performing a simple object naming task had spent just as long gazing at the object as they did when they named it fluently. Apparently, then, disfluency did not result from either long or hasty examination of the object to be named. Disfluency does not appear to be a measure of perceptual problems per se.

Instead, disfluency is related to the cognitive burdens of production [5]. We will use disfluency to discover whether there is a cognitive cost involved in taking up information needed to pursue a dialogue task. We will then show that this cost is put to good use: the locations of disfluencies reveal that they are appropriate responses to the information that speakers have garnered.

The information in question underpins what is thought to be a crucial task in dialogue: each participant must maintain a model of her interlocutors' knowledge so as to adjust to their mutual knowledge both what she says and how she says it. Most views of dialogue now assume that speakers will take some interest in indications both of the listener's knowledge about the domain under discussion and of the listener's satisfaction with the communication just made. Clark and Krych [9], for example, propose that speakers monitor listeners' faces for all manner of feedback, much as they track listeners' utterances. Horton and Gerrig [16] acknowledge the costs of this operation, suggesting that complete uptake and application of listener information could prove to be taxing in some cases, so that utterances will be less perfectly designed for the audience as the cognitive burden increases.

To determine whether garnering cues to listener knowledge is indeed costly to production, we use a variant of the map task [2, 7]. As in the original task, players have before them versions of a cartoon map representing a novel imaginary location. The Instruction Giver communicates to the Instruction Follower a route pre-printed on the Giver's map. The present experiment manipulates time-pressure and the modality or modalities in which a distant confederate delivers pre-scripted feedback to the speaker's instructions. Verbal feedback affirms comprehension of some instructions and declares general incomprehension of others. Visual feedback, in the form of a simulated listener-eyetrack projected onto the map, may correctly go to the named map landmark or wrongly advance to another. Where both modalities are used, their feedback may be concordant or discordant across modalities. Scripted and simulated responses are used to control the conditions under which speakers are operating. Genuine speaker eye-gaze is tracked.

We use eyetracks, rather than sight of the speaker's direction of gaze, to represent listener feedback for two reasons. First, simulated gaze is much easier to control than genuine gaze on the part of the confederate. Second, though facial expressions and direction of gaze have real value, tasks with a visual component produce remarkably little inter-interlocutor gaze [[1,3,11]]. To allow simultaneous performance of the task and uptake of listener information, the listener’s ‘eyetrack’ was superimposed on the map (See Figures 1 and 2).

The present paper will examine two kinds of disfluency diistinguished by previous research, repetitions and deletions. In the current definition, a repetition is produced when the speaker repeats verbatim one or more words with no additions, deletions, or re-ordering, as in (1)

(1) Now you want to ***go* go** just past the tree

Repetitions are thus a single faulty attempt at communicating the same message in the same form. In contrast, a deletion has occurred whent the speaker interrupts an utterance without restarting or substituting syntactically similar elements, as in (2)

(2) A MOVE 36 You need to be just under…
A MOVE 37 Do you have a White Mountain?

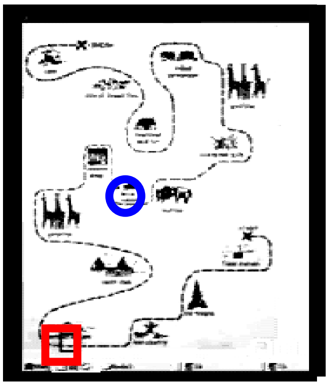
Thus, deletions abandon one communicative act in favour of another.

In this setting, there seem to be two distinguishable predictions. Clark and Krych [9] predict good uptake of all visual cues to listener knowledge and suitable application of the information. Horton and Gerrig [16] predict that the more complex the input, the more difficult will be both uptake of cues and the production of suitable speech. Thus there should in principal be an increase in dsfluency if speakers observe negative visual feedback (’follower gaze’ at wrong landmarks) and if there ar conflicts between verbal and visual feedback.

1.1. Task and procedure

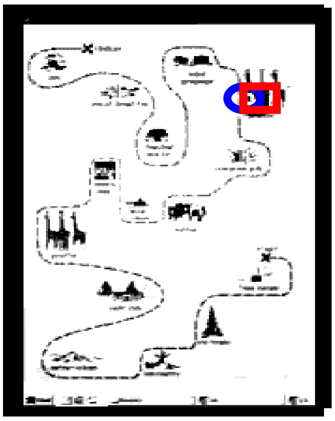
All the materials come from an experiment which used conversations between subject Instruction Givers and a confederate Instruction Follower. Each subject was greeted individually with the confederate. Each subject was naïve to the status of the confederate and during post-experimental debriefing, none reported any suspicions. Both subject and confederate were told that whoever took the role of Instruction Giver should guide the Instruction Follower, from a marked start-point to buried treasure. Subject and confederate then ‘negotiated’ that the subject would be Giver and the two were taken to separate rooms. The Giver was seated 60 cm from a flat screen monitor displaying the map. Labelled landmarks and map designs were adapted from the HCRC Map Task Corpus [2]. Eye tracking movements were recorded using a non-invasive Senso-Motor Instruments remote eye-tracking device placed on a table below the monitor. Eye movements were captured with Iview version 2 software. The tracker was re-calibrated at the beginning of each trial. Speech was recorded in mono using Asden HS35s headphone- microphone combination headsets. Video signals from the eye tracker and the participant monitor were combined and recorded in Mpeg with Broadway Pro version 4.0 software.

Feedback from the confederate took two forms. Visual feedback consisted of a simulated eyetrack, a small red square advancing from landmark to landmark once each landmark was named, and showing saccades of random length and direction. The visual feedback was under the control of the experimenter, who advanced the feedback square to its next programmed position when the Giver first mentioned a new route-critical a landmark. When feedback was scheduled to be wrong, the square moved to a landmark that had not been named. When feedback was to be correct, the feedback square advanced to the landmark just named. Similarly, verbal feedback came from the confederate subject who read pre-scripted responses. Just as with the visual feedback, the confederate provided verbal feedback when the speaker uttered the first mention of the landmark in question. Figures 1 and 2 illustrate possible events.



Instruction Follower:
‘Yes, got it.’

Figure 1. Discordant feedback. Circle = Giver’s gaze; Square = Follower’s feedback (wrong location).



Instruction Follower:
’Okay, that’s fine’

Figure 2. Concordant feedback. Circle = Giver’s gaze; Square = Follower’s feedback (correct location).

1.2. Experimental Design

The experiment crossed feedback modality (3), single modality group (2), and time-pressure (2). In the *No Feedback* conditions, subjects saw only the map. In the *Single-Modality* condition, subjects in the Verbal Group got verbal feedback only, while those in the Visual Group had only visual feedback. Finally, in the *Dual-Modality* condition, all subjects received both visual and verbal feedback. The two modalities might be discordant or concordant. *Concordan*t feedback consisted on average of 8 instances of positive verbal and correct visual feedback, and 6 instances of negative verbal and wrong visual feedback per map. In each map, *discordant* feedback included roughly 3 instances of negative verbal and correct visual feedback, and 6 instances of positive verbal and wrong visual feedback. This design is portrayed in Table 1. In half of the trials, speakers under *time-pressure* had three minutes to complete the task; in *untimed* dialogues there was no time limit.

Table 1. The relationship between the Experimental Groups and the various Feedback Modalities.

Experiment	Feedback Modalities		
	None	Single	Dual
Verbal Group	None	**Verbal**	Verbal + Visual
Visual Group	None	**Visual**	Verbal + Visual

Thirty-six subjects with normal uncorrected vision were recruited from the Glasgow University community. All were paid for their time. All encountered all 6 conditions. Six different basic maps were used, counter-balanced across conditions over the whole design. Subjects were eliminated if any single map trial failed to meet criteria for feedback or capture quality. The feedback criterion demanded that the experimenter advance the feedback square between the introduction of the pertinent landmark and the onset of the following instruction in all cases where the feedback was scheduled to be errant and in 70% where the square's movement was scheduled to be correct. The capture criterion demanded that at least 80% of the eye-tracking data was intact. Subjects were also eliminated if on debriefing they revealed any suspicions about the nature of the interlocutor.

2. Results

2.1. Baseline effects: Words

Since the opportunities for disfluency increase with increasing amount of speech, it is important to note effects of the experiment's design on word counts. Word counts for whole and part-words show less speech with time-pressure (425 words/trial on average) than without (579): (F_1(1,34) = 24.38, p < .001). Visual Group Single-Modality trials (459 words) were shorter than the corresponding Dual-Modality trials (590 words) with no corresponding change for Verbal subjects (Feedback Modality x Group: (F_1(2,68) = 8.65 p < .001; Bonferroni: t = -6.4, p < *.001*). Since Dual-Modality Conditions do not differ between groups (Verbal: 616, Visual: 590), we can use this condition to examine the relationships between disfluency and gaze or dialogue events.

We also examined speech rate across the experimental conditions. To calculate speech rate we divided the Giver words per map by the total Giver speaking time for the map (the summed durations of all conversational moves less the summed durations of both simple and filled pauses). Time-pressure had no significant effect on speech rate. The interaction between Feedback Modality and Group (F_1(2,68) = 4.87, p < .02) presented in Table 2, is due only to a difference between the No-Feedback (.34) and Dual-Modality (.30) conditions for the Verbal Group (Bonferroni p = .004). Again Dual Modality conditions are alike.

Table 2. Speech rate (Words/Total speaking time) means from Feedback Modality x Group interaction

Experiment	Feedback Modalities		
	None	Single	Dual
Verbal Group	.340	.303	.304
Visual Group	.344	.343	.340

2.2. Baseline effects: Gaze

In order to test for the relationship between disfluency and Giver gaze, it was necessary to determine whether all conditions in which a Giver might gaze at a feedback square actually did succeed in directing the Giver's attention to the square. To check for overlap of gaze between Giver and 'Follower', the video record of feedback and Giver Gaze were analyzed frame by frame for the landmark at which each was directed. When Follower Gaze and Giver Gaze were on the same landmark, the Giver was considered to be looking at the feedback square. Here we report the number of feedback episodes [task sub-portions containing in feedback] in which *any* frame contained an instance of gaze at the feedback square].

Givers did not make use of all their opportunities by any means (Figure 3). Nor did they use their opportunities equally (Visual feedback x Verbal feedback: F_1(1,34) = 7.70, p < .01). Strangely enough, Givers used fewest opportunities in an important concordant condition, the one in which the Follower was clearly lost: the Follower square was hovering over a wrong landmark while the Follower was simultaneously providing negative verbal feedback (verbal- vis-: .366). These attracted less gaze than another concordant condition – when the Follower needed no help because she was in the right place and said so (verbal+ vis+: .511). Similarly Givers looked less when the Follower was lost but claimed not to be (verbal+ vis-: .448) than when she was correct but claimed to be lost (verbal- vis+:.591) (Bonferroni t-tests at .008). A simple description says that speakers are most likely to track listeners, the listener's location falls under their own gaze, which is occupied by the things they are describing. Apparently, spekaers prefer not to go off-route to learn the whereabouts of an errant follower.

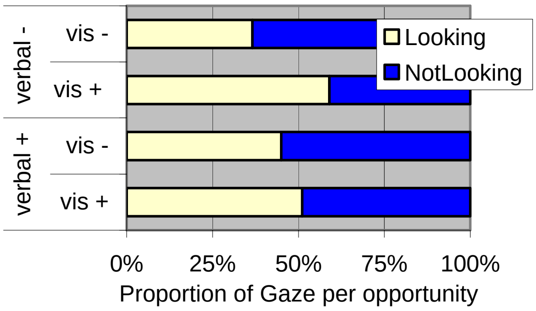


FIGURE. 3 Proportion of feedback episodes attracting speaker gaze to feedback square: Effects of combinations of visual and verbal feedback in dual channel conditions

2.3. Disfluencies Overall

The first author labeled disfluencies according to the system devised by Lickley [19] as repetitions, insertions, substitutions or deletions. She used Entropic/Xwaves software to listen to, view and label disfluent regions of speech. Spectrograms were analyzed whenever necessary. Each word within a disfluent utterance was labeled as belonging to the reparandum, the interregnum, or the repair. A reparandum involves speech that is either overwritten, expunged or retraced in the repair [19]. Repairs typically 'replace' the error in the reparandum. Since deletions are typically abandoned utterances, they have no repair [19, 27].

Because disfluencies are more common in longer utterances [6, 10, 25] we divided the number of disfluencies in a monologue by its total number of words, yielding disfluency rate as a dependent variable.

Disfluency rates were submitted to a by-subjects ANOVA for Group (2) (Verbal vs. Visual), Time-pressure (2) (timed vs. untimed) and Feedback Modality (3) (none, Single-Modality, Dual-Modality). The baseline No-Feedback conditions differed between Verbal and Visual groups (Group * Modality: F_2(2,68) = 5.21, p < .01; Bonferroni, t = 2.94, p < .02). This difference can be explained by a single subject in the Verbal Group who was an outlier in terms of disfluency. Because of this subject, there was no effect of Feedback Modality within the Verbal Group, while the Visual Group showed the expected increase in rate of disfluency between No Feedback and Single- (Bonferroni t = -4.12, p = .001) or Dual-Modality conditions (Bonferroni t = -5.77, p < .001). Since Single and Dual Modality conditions did not differ, we can proceed to examine only the Dual Modality conditions in the expectation that conflicting feedback (only found in Dual Modality) *per se* is not an overall cause of disfluency.

2.4. Disfluency Types: Repetitions v Deletions

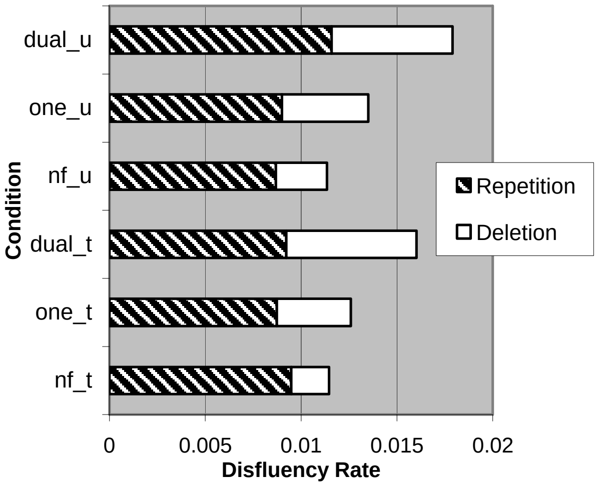


Figure 4. Rates of disfluency by type and experimental condition for the Verbal and Visual Groups combined. nf = no feedback, one = Single-Modality feedback, dual = Dual modality feedback; t = timed, u = untimed.

An initial investigation of deletions and repetitions begins to separate them. Figure 4 displays their distributions across experimental conditions. Independent analyses were done for each type of disfluency; that is one analysis within deletions only and one within repetitions only.

As found in [23], only deletion rate showed any significant effect of feedback: Deletion rate rose significantly with each additional feedback modality (No Feedback .002, Single-Modality .004, Dual-Modality .007; $F_1(2,68) = 21.00$, $p < .001$; all Bonferroni t-values < .01). There were no effects of time-pressure on deletion rate and no significant interactions.

For repetitions on the other hand, an interaction between Time-pressure and Group ($F(1,34) = 6.27$, $p < .02$) revealed that subjects were more disfluent in the untimed condition (.012) of the Verbal Group than they were anywhere else in either the Verbal or the Visual Group, timed or untimed, though the internal comparisons were not significant.

3.5 Disfluency & Eye-Gaze

Within the Dual-Modality condition, the experimental design contrasted positive and negative feedback in the two modalities. However, the modalities are concordant or discordant only if the Giver actually takes up both visual and verbal feedback. The tendency for more speech in conditions with verbal feedback suggests that subjects were attending to what the confederate Follower said. Eye-tracking enabled us to tell when the Giver had actually looked at the Follower's visual feedback. As Figure 3 made plain, Givers do not take up the same proportion of concordant and discordant feedback. They gazed most at one kind of discordant feedback (negative verbal + correct visual) and least at a concordant condition (negative + wrong visual feedback).

To look for disfluency in truly vs potentially concordant and discordant situations, we examined disfluency per feedback opportunites in concordant and discordant situations contrasting those in which Givers did or did not look at Follower feedback. In fact, Givers who attended to discordant feedback from the Follower encountered subsequent fluency problems. The number of disfluencies per feedback opportunity was greatest following a discordant feedback episode in which the Giver had actually gazed at the Follower feedback square (.333), a significantly higher rate than following a concordant feedback episode which had drawn the Giver's attention (.205) (Bonferroni $t = -3.51$, $p = .001$ within by-subjects Group (2) x Giver attention (looking v not looking) x Concordance of modalities (concordant v discordant: $F_1(1,34) = 7.24$, $p = .01$). None of the other pairwise comparisons was significant.

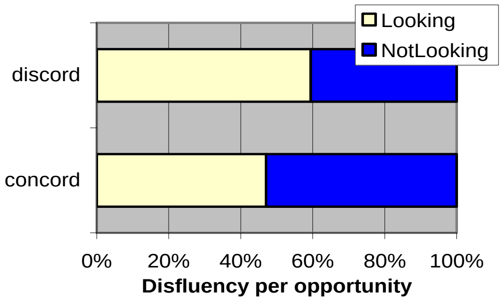


Figure 5. Rate of repair disfluencies per concordant or discordant feedback opportunity with respect to whether the Giver was either looking or not looking at the Follower. The difference is significant when the Giver looked at the Follower.

3.6 Disfluency Type, Gaze & Motivation

So far we have seen that speakers' gaze behaviour is not randomly distributed. It follows certainly problems (a Follower on-route who claims not to be) and ignores others (a Follower off-route who claims to be on-route). We have also seen that on those occasions when an instruction Giver actually takes in enough information to see what is amiss, he or she is more likely to speak disfluently. The question we ask here is whether these disfluencies are part of well formed communicative processes. If the information taken in by examination of listener feedback is properly processed by the speaker, what s/he says disfluently will be something appropriate to the situation. To determine whether this is really the case, it was necessary to classify utterances by their goal or motivation. To do this, the first author examined all 564 repetitions and 280 deletions occurring in the Dual Modality feedback condition.

The first stage of this process was to identify an interval for analysis. All dialogues were coded according to the HCRC Conversational-Game-Move coding scheme [8]. In this system, each turn is decomposable into conversational Moves, or sub-units of the dialogue. For example, a speaker might 'Instruct' by giving directions or 'Align' when noting that the Follower has gone astray. Analyses began with the Move that carried the disfluency. The coder searched backwards from the Interruption Point of the disfluency to the most recent Giver Move introducing a new landmark. The start time was considered to be the Giver's first mention of a new landmark while the end time was the Interruption point of the disfluency or for deletions, the end of the repair.

The second stage was to identify Giver gaze behaviours within these intervals. The gaze record of the speaker for this time-span was then checked and disfluency was coded as 'Looking' if there were any overlaps of Giver and Follower Gaze from the introduction of the landmark to the end of the disfluency. All others were coded 'Not Looking'.

Third, each disfluency was classified by Motivation, the content of the repair. Repetitions necessarily occur within the same dialogue Move, while deletions are almost always a single abandoned Move, so that the repair effectively lies in the next Move. Motivations were classified under two major goals: either the speaker was ‘confirming’ that the Follower was at a correct or incorrect landmark or the speaker was ‘reformulating’ by adding, elaborating, or correcting information being transmitted. Examples of goal and disfluency combinations are given in Table 3 below.

Table 3. Examples of disfluencies by goal and type. For repetitions, both reparandum and repair appear in bold text. For deletions, just the reparandum appears in bold text since the repair is effectively non-existent.

Disfluency Type	Dialogue Goal	
	Confirmation	Reformulation
Repetition	‘**That’s, That’s** just fine	‘Eh you travel directly ehm sort of **north…north** and east’
Deletion	‘So loop around the waterfall **over**….Yeah, there’	‘Um **can you si-**…it’s to the left of that’

Since appropriate confirmation of position should depend on the Giver actually determining where the Follower was, we would expect confirmations to accompany gaze at the follower. Since the arrival of the Follower at the goal or her movement off route should complete the execution of a series of instructions, all the Giver need do is cease instructing and declare the Follower to be right or wrong. Accordingly, deletion disfluencies are appropriate: in this view they mark a sequence of instructing, checking, and, finally, abandoning any ongoing instruction for a new a phase in the dialogue.

Our second goal category, reformulation, can also repair communication problems but by elaborating the material serving the current goal. Typically [14], speakers have to look away from their interlocutors when formulating complex material. Also on the grounds of complexity, we might expect not looking and reformulating to accompany repetition disfluencies [10].

Analyses of Giver’s Gaze (2: looking vs. not looking), Motivation (2: confirmation vs. reformulation), Disfluency Type (2: repetition vs. deletion) and Time-pressure (2: timed vs. untimed) showed part of this pattern.

We predicted that reformulations would attract repetition disfluencies and confirmations would attract deletions. As Figure 6 illustrates, numerically repetitions (confirmation = 0.083; reformulation = 0.403) and deletions (confirmation = 0.245; reformulation = 0.186) worked as predicted ($F_1(1,34)$ = 59.60, p < .001). The predicted effect of Motivation, however, was significant only for repetitions (F_1 (1,34) = 124.17, p < .001).

We predicted that looking at the feedback square would yield confirmations and not looking would accompany reformulations. In fact, only when Givers did not gaze at the Follower’s square was the prediction met: there was a higher rate of reformulations than confirmations (Gaze x Motivation: *F(1,34) = 9.27, p < .01,* Bonferroni *t* at *p* = .008.).

Since we have an association between reformulations and repetitions, and one just reported between reformulations and not looking at the interlocutor, we tested for the effects within repetitions and deletions separately. Though the Giver tended not to look at the Follower square during repetition disfluencies, the trend is weak because it appears to hold only in the Verbal Group (Disfluency Type x Gaze: *F(1,34) = 3.59, p = .067;* Gaze x Motivation x Experiment: *F(1,34) = 8.62, p < .006;* Bonferroni at *p* = .001). For deletion disfluencies, the effect of gaze depends on motivation: deletions classified as confirmations were, as we predicted, more common when the Giver took the opportunity to look at the Follower (Bonferroni at *p* = .008), whereas deletions classed as reformulations showed an insignificant tendency to be more common when the Giver was not looking at the Follower (Motivation x Gaze: *F(1,34) = 8.61, p < .01).* Thus, there were associations between disfluency type and motivation type and between disfluency-motivation combination and gaze.

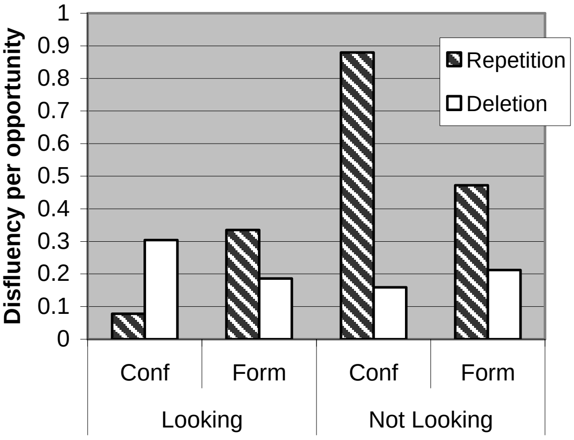


Figure 6. Rates of Repetitions and Deletions per opportunity with respect to Behaviour type, either confirmation (Conf) or reformulation (Form) and Gaze. The difference is significant for Repetititions but not for Deletions.

3. Discussion and Conclusions

Although the visual feedback provided the Giver with the Follower’s exact location at any point during the interaction, this information had a cost. The Giver tended to gaze away from the Follower’s location. Gaze aversion during difficulty is a common phenomenon found in conversational analysis and gaze studies [14, 15], and we find that gaze itself makes for production difficulty: speakers are more disfluent if they look at the follower feedback. Furthermore, Givers tended not to look at concordant negative feedback which clearly indicated trouble, though they did look at discordant feedback when the Follower was easily found – on the landmark being described.

When a Giver noticed this discordance, disfluency often occurred as result, presumably because the speaker was burdened with resolving the conflicting verbal and visual signals and in a sense handling the Follower’s confusion. Disfluency, it seems, tend to co-occur first with uptake of the speaker’s whereabouts and misalignment in dialogue, as predicted in [24]

If speakers are committed to tracking and accommodating listeners’ knowledge [9, 10], and if repetitions indicate commitment to listener and message, Givers should visually attend to their Followers whilst making a repair: a committed speaker might be expected to assist a Follower who is clearly in difficulty by looking at the Follower’s feedback and tailoring any following utterances to them. Instead, repetitions tended to associate with reformulation and thus by reformulation to gaze aversion during critical need. Looking at the follower instead accompanied deletions, as the Giver abandoned a Move in order to confirm or deny the listener’s progress. Thus, it seems deletions, or false starts were associated with attending to the Follower but not with commitment to the utterance.

The present paper has added a psycholinguistic and dialogue perspective to the taxonomy of disfluency. We found that speakers are disfluent in different ways depending upon the dialogue task in which they are currently engaged. The nature of listener feedback and the Giver's uptake of information about the listener both had effects.

4. Acknowledgements

The authors thank Yiya Chen and Catriona Havard for running the experiment and administering the gaze and transcription coding. This work was supported by EPSRC Research Grant GR/R59038/01 to Ellen Gurman Bard and GR/R59021/01 to Anne H. Anderson.

5. References

[1] Anderson, Anne H., Ellen Gurman Bard, Cathy Sotillo, Alison Newlands, & Gwyneth Doherty-Sneddon. 1997. Limited Visual Control of the Intelligibility of Speech in Face-to-Face Dialogue. *Perception and Psychophysics,* 59 (4), pp. 580-592.

[2] Anderson, Anne H., Miles Bader, Ellen Gurman Bard, Gwyneth Doherty, Simon Garrod, Steve Isard, Jacqueline Kowtko, Jan McAllister, Jim Miller, Cathy Sotillo, Henry S. Thompson, & Regina Weinert, 1991. The HCRC Map Task Corpus. *Language and Speech***,** vol. 34, pp. 352–366.

[3] Argyle, Michael & R Ingham. 1972. Gaze, mutual gaze and proximity. *Semiotica*, 6. pp. 289-304.

[4] Bard, Ellen Gurman, Anne H. Anderson, Marisa Flecha-Garcia, David Kenicer, Jim Mullin, Hannele Nicholson, Lucy Smallwood & Yiya Chen, 2003. Controlling Structure and Attention in Dialogue: The Interlocutor vs. the Clock. *Proceedings of ESCOP, 2003*, Granada, Spain.

[5] Bard, Ellen Gurman, Robin J. Lickley, & Matthew P. Aylett. 2001. Is Disfluency just Difficulty? *Proceedings of DiSS'01*, Edinburgh.

[6] Bard, Ellen Gurman, Anne H. Anderson, Cathy Sotillo, Matthew Aylett, Gwyneth Doherty-Sneddon & Alison Newlands. 2000. Controlling the intelligibility of referring expressions in dialogue. *Journal of Memory and Language*, vol. 42, pp. 1–22.

[7] Brown, Gillian, Anne H. Anderson, George Yule, Richard Shillcock, 1983. *Teaching Talk*. Cambridge: Cambridge University Press.

[8] Carletta, Jean, Amy Isard, Steve Isard, Jacqueline Kowtko, Gwyneth Doherty-Sneddon, & Anne H. Anderson, 1997. The reliability of dialogue structure coding scheme. *Computational Linguistics*, vol. 23, pp. 13–31.

[9] Clark, Herbert H. & Meredyth A Krych. 2004. Speaking while monitoring addresses for understanding. *Journal of Memory and Language.* vol. 50, Issue 1, pp. 62-81.

[10] Clark, Herbert H. & Thomas Wasow, 1998. Repeating words in Spontaneous Speech. *Cognitive Psychology*, vol. 37, pp. 201–242.

[11] Exline, Ralph V., P. Jones, & K. Maciorowski. 1977. *Race, affiliation-conflict theory and mutual vision attention during conversation.* Paper presented at the meeting of the American Psychological Association.

[12] Fox Tree, Jean & Clark, Herbert H.. 1997. Pronouncing 'the' as 'thee' to signal problems in speaking. *Cognition. 62. pp. 151-167*

[13] Fox Tree, Jean. 1995. The effects of false-starts and repetitions on the processing of subsequent words in spontaneous speech. *Journal of Memory & Language.* 34. pp. 709-738.

[14] Glenberg, Arthur M, Jennifer L. Schroeder & David A. Robertson. 1998. Averting the gaze disengages the environment and facilitates remembering. *Memory and Cognition.* Vol. 26, (4). pp. 651-658

[15] Griffin, Zenzi M., 2005. The Eyes are right when the Mouth is Wrong. *Psychological Science*. Vol 15, number 12, pp. 814-821

[16] Horton, William S. & Richard J. Gerrig. 2005. The impact of memory demands on audience design during language production. *Cognition,* vol. 96. pp. 127-142.

[17] Levelt, Willem J.M., 1989. Monitoring and self-repair in speech, *Cognition,* vol. 14, pp. 14–104.

[18] Lickley, Robin J. 2001. Dialogue Moves and Disfluency Rates. *Proceedings of DiSS '01, ISCA Tutorial and Workshop*, University of Edinburgh, Scotland, UK, pp. 93-96.

[19] Lickley, Robin J. 1998. HCRC Disfluency Coding Manual *HCRC Technical Report* 100. `http://www.ling.ed.ac.uk/~robin/maptask/disfluency-coding.html`

[20] Lickely, Robin J. 1995. Missing Disfluencies. *Proceedings of ICPhS,* Stockholm, vol. 4. pp. 192-195.

[21] Lickley, Robin J. 1994. *Detecting Disfluency in Spontaneous Speech.* PhD. Thesis, University of Edinburgh.

[22] Maclay, Howard & Charles E. Osgood. 1959. Hesitation phenomena in spontaneous English speech. *Word,* 15, pp. 19-44.

[23] Nicholson, Hannele, Ellen Gurman Bard, Robin Lickley, Anne H. Anderson, Jim Mullin, David Kenicer & Lucy Smallwood, 2003. The Intentionality of Disfluency: Findings from Feedback and Timing. *Proc. Of DiSS'03, Gothenburg Papers in Theoretical Linguistics 89. pp.15-18*

[24] Pickering, Martin & Simon Garrod, 2004, Towards a mechanistic theory of dialogue: The interactive alignment model. *Behavioral & Brain Sciences*. 27 (2), pp. 169-190.

[25] Plauché, Madelaine & Elizabeth Shriberg, 1999. Data-Driven Subclassification of Disfluent Repetitions Based on Prosodic Features. *Proceedings of the International Congress of Phonetic Sciences,* vol. 2, pp. 1513–1516, San Francisco.

[26] Savova, Guergana & Joan Bachenko. 2002. Prosodic features of four types of disfluencies. *Proceedings of DiSS'03.* Gothenburg University, Sweden. pp. 91-94.

[27] Shriberg, Elizabeth. 1994. Preliminaries to a Theory of Speech Disfluencies. PhD Thesis. University of California at Berkeley.

[28] Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, Julie Sedivy. 2000. Integration of visual and linguistic information in spoken language comprehension. *Science.* 268, pp. 1632-1634.