Disfluencies in dialogues with patients with schizophrenia

Christine Howes (christine.howes@gu.se)
Philosophy, Linguistics and Theory of Science
University of Gothenburg, Sweden

Mary Lavelle
Florence Nightingale School of Nursing and Midwifery
King's College London, UK

Patrick G.T. Healey
Queen Mary University of London

Julian Hough
Bielefeld University

Rose McCabe
Exeter University

Abstract

Disfluencies such as self-repairs, filled pauses such as 'um' and silent pauses are pervasive in dialogue, but there is no consensus in the literature as to whether they reflect internal production pressures, or interactive issues – or how their effects are manifest in dialogue. It is well-known that patients with schizophrenia have problems with language and social cognitive skills, yet little research has investigated how these impact interaction. We report a study on the disfluency behaviours of patients with schizophrenia and their interlocutors who were unaware of the patient's diagnosis, compared to healthy control groups. Results show that patients use fewer self-repairs than either their partners or controls and fewer filled pauses ('er', 'um') than controls. Furthermore, the presence of the patient also affects patients' partners, who use fewer filled pauses than controls and more unfilled pauses than both patients and controls. This suggests that smooth coordination of turns is problematic in patients' dialogues.

Keywords: Disfluency; Dialogue; Schizophrenia

Introduction

Disfluencies, such as self-repair, pauses and filler non-words such as *er* and *um* (*filled pauses*) are pervasive in dialogue (Schegloff, Jefferson, & Sacks, 1977). Such disfluencies are conventionally regarded as symptomatic of problems with communication, caused by self-monitoring or production issues (Levelt, 1983). However, disfluencies also highlight the interactive nature of dialogue – many disfluencies occur as we tailor our talk for specific addressees, or as a direct result of feedback from our interlocutors (Goodwin, 1979).

Furthermore, different types of disfluencies have been hypothesised to contribute differently to the individual and shared actions that must be coordinated in successful dialogue. For example, in route following experiments, different distributions of filled pauses and self-repairs suggest that self-repairs occur because of production difficulties, but filled pauses fulfil an interpersonal function (Bortfeld, Leon, Bloom, Schober, & Brennan, 2001; Nicholson, Eberhard, & Scheutz, 2010; Brennan & Schober, 2001).

Other research suggests that disfluencies should be categorised according to whether they make changes to the meaning of an utterance or not, e.g. reformulations and false starts are *backwards-looking* disfluencies, whilst word repetitions and filled pauses are *forwards-looking* (Ginzburg, Fernández, & Schlangen, 2014; Allwood, Nivre, & Ahlsén, 1990).

In addition to signalling difficulties that the speaker may be experiencing, different types of disfluency have been shown to have conventionalised meanings with respect to turn-taking. Filled pauses may indicate a break in the information (for example while the speaker searches for a word or phrase) but an intention to retain the floor, whilst unfilled pauses may signal that the speaker does not intend to continue speaking (Clark & Fox Tree, 2002; Allwood et al., 1990).

Recent work suggests that disfluencies have measurable effects on the dialogue – even if this is not necessarily the intention of the speaker (Finlayson & Corley, 2012; Ginzburg et al., 2014). For example, in contexts where informational exchange is key, such as the Map Task (Anderson et al., 1991), more self-repair may be indexing one's own production difficulties, and how hard people are working to be understood by (and for) their interlocutors (Colman & Healey, 2011).

Further evidence that disfluencies do not just function as markers of miscommunication but contribute to improving the effectiveness of interaction comes from psycholinguistic studies. For example, referential success and ambiguity resolution are aided by the presence of disfluencies (Brennan & Schober, 2001; Bailey & Ferreira, 2007). Communications training interventions also indicate that talk between psychiatrists and patients with schizophrenia is improved when the psychiatrist uses more self-repair (McCabe et al., 2016).

As can be seen from the above discussion, there is no general consensus in the literature about either which disfluencies should be focused on, how different types of disfluency should be categorised, or the effects they have in dialogue.

It is well documented that people with a diagnosis of schizophrenia have problems with language and social cognitive skills, including with self-monitoring (Johns et al., 2001) and turn-taking (using role-play; Mueser, Bellack, Douglas, and Morrison, 1991), yet little research has investigated how these impact interaction. The few studies that do find that less patient self-repair is associated with verbal hallucinations (Leudar, Thomas, & Johnston, 1992), more patient other-initiated repair (clarification of the doctor's talk) is associated with better adherence to treatment (McCabe et al., 2013), and clinicians' use of self-repair has positive clinical consequences (McCabe et al., 2016). Research into disfluencies therefore has the potential to be used in diagnostic tools, and feed into training for psychiatrists to detect when a patient is in difficulty or shape their own talk more effectively.

However, there is conflicting evidence regarding disfluencies in patients with schizophrenia and whether this differs from non-clinical populations. In consultations, patients use more self-repair than psychiatrists (McCabe et al., 2013) and this is higher than found in general dialogue in the demographic portion of the British National Corpus, or Map Task dialogues (Howes, Purver, McCabe, Healey, & Lavelle, 2012;

Colman & Healey, 2011). This might be expected in the clinical domain where patients are explaining and providing updates on their health and treatment. In contrast, in a controlled study where subjects described the experimenters actions, frequency of repair did not differ between patients and matched controls (Leudar et al., 1992).

These differences may reflect interactional factors, such as domain, or role, and not differences between schizophrenia patients and non-clinical populations per se. Additionally, in these studies, patients' interlocutors were aware of their diagnosis, which could have influenced the way they interacted with the patient (Doyen, Klein, Pichon, & Cleeremans, 2012).

We report an analysis of a unique corpus of 40 triadic dialogues (Lavelle, Healey, & McCabe, 2013) that avoids these potential confounds (see Method, below). Given that the literature suggests patients produce more self-repairs than their psychiatrists in dialogues during clinical consultation, and more so than people in natural corpora from the general population, and with the assumption that turn-taking cues (including offering the floor to another participant, or retaining it for oneself) partially consist of filled and unfilled pauses (Maclay & Osgood, 1959), we expect the following:

Hypotheses

1. Compared to healthy control conversational groups and their healthy conversational partners, patients will produce more disfluent talk, with more self-repairs.

2. Compared to controls and their conversational partners, patients will produce fewer turn-taking cues (filled and unfilled pauses).

3. Compared to patients and controls, patients' partners will produce more turn-taking cues.

Method

Participants

The data consist of transcripts of twenty patient interactions, involving one patient conversing with two healthy controls who were unaware of the patient's diagnosis, and twenty control interactions (3 healthy participants). Twenty patients with a diagnosis of schizophrenia (6 male, 14 female) and one hundred non-psychiatric healthy participants, forty in the patient condition (21 m, 19 f) and sixty in the control condition (34 m, 26 f), participated. Participants within each triad were unfamiliar to each other. Due to technical issues one patient interaction and one control conversation could not be transcribed and are excluded from the analysis, resulting in data from 57 individuals in control groups (19 triads), and 19 patients and 38 healthy controls in patient interactions.

Non-psychiatric healthy participants were recruited through advertising on local community websites. Of those who responded to the advertisement, 40% participated. Participants with a diagnosis of psychosis or affective disorders in themselves, or any first-degree relatives, and those who were not fluent English speakers were excluded.

Patients were recruited at routine psychiatric outpatient clinics under supervision of their psychiatrist. 25% of all patients approached agreed to participate. Patients were taking anti-psychotic medication which fell within the low dose range (Chlorpromazine equivalents 50-200mg/day). Non-native English speakers and patients presenting with motor side effects from antipsychotic medication were excluded based on a clinician's assessment.

The distribution of gender did not significantly differ between patient and control conditions (P: $n = 60$: female $= 53.33\%$, C: $n = 60$: female $= 43.33\%$; $\chi^2 = 1.20, p = 0.27$). Patients were significantly older than controls (P: $M = 41 s.d. = 8.6$, C: $M = 31 s.d. = 9.6$; $t_{119} = 4.51, p < 0.01$)

Symptoms were assessed using the Positive And Negative Symptom Scale for Schizophrenia (Kay, Flszbein, & Opfer, 1987). Patients displayed relatively low PANSS scores for positive symptoms – additional features that occur with the disorder such as hallucinations or delusional beliefs ($M = 15.8; s.d. = 6.76$), and negative symptoms, which represent a reduction in usual function such as social withdrawal, diminished affect, apathy and anhedonia ($M = 9.95; s.d. = 3.36$).

Ethics

All procedures were approved by a UK NHS Research Ethics Committee (07/H0711/90). All participants gave written informed consent and were free to withdraw at any time.

Procedure

Participants were brought into the laboratory in threes and seated in a triangular formation so that they all had good visual access to each other (see Figure 1). The researcher read aloud a fictional moral dilemma, the 'balloon task' (see Task section, below), which has been used for studying dialogue, and is known to stimulate discussion (Howes, Purver, Healey, Mills, & Gregoromichelaki, 2011). 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 agreement had their interaction terminated at approximately 450 seconds (7 minutes 30 seconds).

Figure 1: Participants engaged in triadic interaction

Task

The *balloon task* is an ethical dilemma requiring agreement on which of four passengers should be thrown out of a hot air balloon, which is losing height and about to crash into some mountains killing all on board unless one of them jumps to their certain death in order to save the other three. The four passengers are described to the participants as follows:

Dr. Robert Lewis – a cancer research scientist, who believes he is on the brink of discovering a cure for most common types of cancer.

Mrs. Susanne Harris – who is not only widely tipped as the first female MP for her area, but is also over the moon because she is 7 months pregnant with her second child.

Mr. William Harris – husband of Susanne, who he loves very much, is the pilot of the balloon, and the only one on board with balloon flying experience.

Miss Heather Sloan – a 9 year-old music prodigy, considered by many to be a "twenty first century Mozart".

Participants were instructed to debate the reasons for and against each person being saved, and reach mutual agreement about who should jump.

Analysis

Participants' speech was transcribed in ELAN (Brugman & Russel, 2004), allowing us to map the transcriptions to the video and precisely time pauses.

Self-repair Self-repairs were annotated automatically using STIR (STrongly Incremental Repair detection; Hough and Purver, 2014);[1] which detects speech repairs on transcripts word-by-word incrementally. It uses a pipeline of classifiers to tag each word of the transcript as either fluent, or in an element of the three-part repair structure below, according to the manual by (Meteer, Taylor, MacIntyre, & Iyer, 1995):

$$\text{John } \underbrace{[\text{ likes }}_{\text{reparandum}} + \underbrace{\{\text{er}\}}_{\text{interregnum}} \underbrace{\text{loves }]}_{\text{repair}} \text{Mary} \qquad (1)$$

STIR uses features from n-gram language models in a pipeline of classifiers which classify whether the current word constitutes a boundary of each part of the repair structure. STIR is trained on the Switchboard corpus (Godfrey, Holliman, & McDaniel, 1992) and achieves an F-score accuracy for self-repair detection of 0.81 on conversational data (Howes, Hough, Purver, & McCabe, 2014). It has previously been applied to therapeutic dialogue, with high rates of correlation to human coders in terms of self-repair rate (Howes et al., 2014), so is adequate for our annotation purposes.

Filled pauses In this data, filled pauses were found to be inconsistently spelt (*aammm*, *er*, *eerrrrmm*, *uhmmm* etc). A find-and-replace operation was applied to the corpus prior to analysis to give these a standardised spelling, i.e. 'er' (Howes et al., 2014). For the analyses we used a count of the number of filled pauses used by each participant.

[1] https://bitbucket.org/julianhough/stir

Unfilled pauses Following e.g. Zellner (1994), we defined unfilled pauses as speech-free spaces between segments of speech by the same speaker of greater than 200 milliseconds. Pause segments were automatically extracted from the ELAN transcripts. For the analyses we used a count of the number of unfilled pauses used by each participant.

As patients produce fewer turns than their interlocutors, per-turn rates for these measures were calculated for each individual participant as the total number of self-repairs, filled or unfilled pauses produced divided by the total number of turns. Patients' turns are also typically shorter – as longer turns are expected to have more repair (Bortfeld et al., 2001), we also calculated measures per 100 words (see Table 1).

Statistics Analyses were run in SPSS using Generalised Linear Mixed Models (GLMMs) to control for both fixed and random effects. In all reported models, condition was a fixed effect and participant ID was a random effect with individuals clustered by their conversation group. For each model we used random intercepts and the maximal random effect structure justified by the sample (Barr, Levy, Scheepers, & Tily, 2013), using a gamma distribution with a log link function. We report exact p-values throughout, but take $p < 0.05$ to be the criterion of significance.

Results

Table 1 shows statistically significant main effects of condition. Pairwise comparisons are reported in the text below.

Number of turns Patients (P) produced significantly fewer turns than their partners (PP) ($t_{1,102} = -3.247, p = 0.001$). There was no significant difference between patients and controls (C) or patients' partners and controls (P/C $t_{1,102} = -1.574, p = 0.118$; PP/C $t_{1,102} = -0.213, p = 0.832$).

Number of words Patients produced fewer words in total and per turn than either their partners or controls (Total: P/PP $t_{1,102} = -3.914, p < 0.001$; P/C $t_{1,102} = -2.481, p = 0.015$; PP/C $t_{1,102} = 0.172, p = 0.864$; Per turn: P/PP $t_{1,102} = -3.823, p < 0.001$; P/C $t_{1,102} = -2.183, p = 0.031$; PP/C $t_{1,102} = 0.979, p = 0.330$).

Table 1: Overview.

		Patient	Partner	Control	F	p
total	Turns	43.78	56.22	57.43	6.822	0.001
	Words	247.67	439.03	399.63	10.069	<0.001
	Self-repair	3.563	10.125	10.490	7.825	0.001
	Filled pause	2.111	4.472	8.078	7.825	0.001
	Unfilled pause	14.944	32.917	22.000	7.372	0.001
per turn	Words per turn	5.58	7.99	7.21	7.141	0.001
	Self-repair	0.081	0.186	0.181	10.708	<0.001
	Filled pause	0.046	0.081	0.140	4.338	0.016
	Unfilled pause	0.362	0.650	0.426	5.342	0.006
per 100 words	Self-repair	1.330	2.203	2.507	6.472	0.002
	Filled pause	0.774	1.004	1.811	5.936	0.004
	Unfilled pause	6.582	7.691	5.929	2.038	0.136

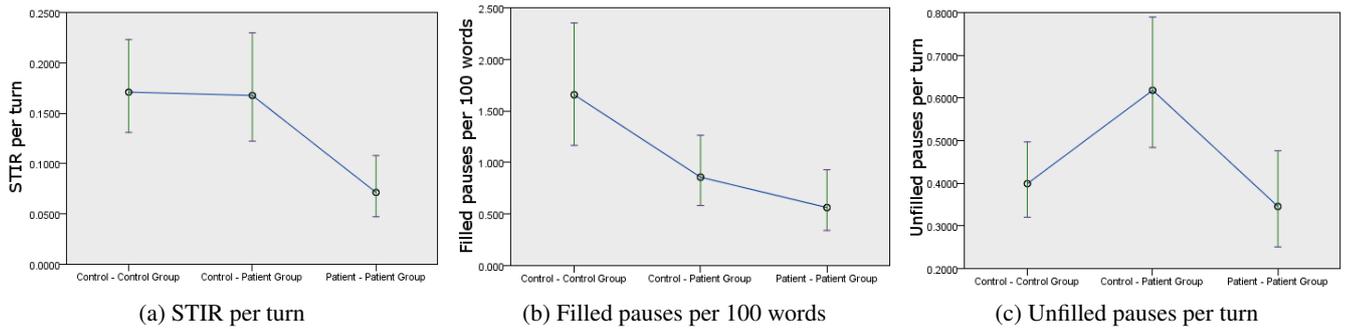| (a) STIR per turn | (b) Filled pauses per 100 words | (c) Unfilled pauses per turn |

Figure 2: Estimated marginal means. Error bars show 95% confidence intervals

Self-repair   For all three measures of repair –total, per turn (shown in Figure 22a) and per 100 words– patients use fewer repairs than either their partners or controls (Total: P/PP $t_{1,96} = -3.265, p = 0.002$, P/C $t_{1,96} = -3.394, p = 0.001$, PP/C $t_{1,96} = -0.528, p = 0.599$; per turn: P/PP $t_{1,96} = -3.476, p = 0.001$, P/C $t_{1,96} = -3.643, p < 0.001$, PP/C $t_{1,96} = -0.094, p = 0.926$; per 100 words: P/PP $t_{1,102} = -2.508, p = 0.014$ P/C $t_{1,102} = -3.280, p = 0.001$, PP/C $t_{1,96} = -0.796, p = 0.428$).

Filled pauses   Patients use fewer filled pauses than controls, with patients' partners levels of filled pauses lying somewhere in between. For the total model, patients' partners use significantly more than patients, and in the per 100 words model (shown in Figure 22b) patients' partners use significantly fewer than controls (Total: P/PP $t_{1,102} = -2.685, p = 0.008$, P/C $t_{1,102} = -3.644, p < 0.001$, PP/C $t_{1,102} = -1.751, p = 0.083$; Per turn: P/PP $t_{1,102} = -1.310, p = 0.193$, P/C $t_{1,102} = -2.720, p = 0.008$, PP/C $t_{1,102} = -1.729, p = 0.087$; per 100 words: P/PP $t_{1,102} = -1.544, p = 0.126$, P/C $t_{1,102} = -3.353, p = 0.001$, PP/C $t_{1,102} = -2.364, p = 0.020$).

Unfilled pauses   Patients' partners use more unfilled pauses in total and per turn (see Figure 22c) than either patients or controls. This is not significant for the per 100 words model (Total: P/PP $t_{1,102} = -3.750, p < 0.001$, P/C $t_{1,102} = -1.996, p = 0.049$, PP/C $t_{1,102} = 2.361, p = 0.020$; Per turn: P/PP $t_{1,102} = -3.235, p = 0.002$, P/C $t_{1,102} = -0.750, p = 0.455$, PP/C $t_{1,102} = 2.483, p = 0.015$, per 100 words: P/PP $t_{1,102} = -1.106, p = 0.272$, P/C $t_{1,102} = 0.644, p = 0.521$, PP/C $t_{1,102} = 1.926, p = 0.057$)

Discussion

The results show differences in disfluencies between the groups, such that patients use fewer self-repairs than either their partners or controls and fewer filled pauses ('er') than controls. Furthermore, the presence of the patient also affects patients' partners, who use fewer filled pauses than controls and more unfilled pauses than both patients and controls.

These results take a coarse-grained view of the data at the level of the individual, and the groups are unbalanced (with 57 controls, 19 patients and 38 patients' partners) which necessarily means that any interpretation is suggestive rather than definitive. However, despite these caveats, we see marked differences between the ways in which the different groups use the different types of disfluency. This suggests several avenues for research which takes a finer-grained approach, by looking at the interactions at the level of the utterance.

For self-repairs, patients produce fewer than either their partners or controls, which is contrary to our expectation (Hypothesis 1) given that patients are known to produce more self-repairs than their psychiatrists in clinical consultations, and than in the demographic portion of the British National Corpus. This may be due to context – patients' engagement is likely to be higher in their psychiatric consultations, compared to first meetings with unfamiliar individuals discussing an abstract topic. The nature of the task is clearly different, with introspective therapy requiring different contributions from the patients in terms of speech production and planning than rational problem solving. Furthermore, participant roles and task demands are more symmetric in the current task compared to clinical consultations. However, if self-repairs are only due to self-monitoring problems, which patients are known to have difficulties with (Johns et al., 2001) we would still expect consistent self-repair patterns across a range of contexts, which does not appear to be the case.

This suggests that we may need to consider the distinction of backwards and forwards looking repair, as proposed in Ginzburg et al. (2014). If repetitions are more like filled pauses, and function as a turn-holding strategy whilst reformulation repairs are backwards-looking, then we expect a difference in the distributions of the different repair types such that patients use fewer repetitions (in line with their rarer use of filled pauses), and that self-repairs that patients do produce are likely to be reformulations, due to patients' self-monitoring problems. It would also be instructive to see if the distributions in patients' partners and controls self-repair are equivalent. The differences in the filled pause data suggests they might not be, which is an avenue for future research.

Partial support for Hypotheses 2 and 3 comes from the data on pauses, although there is conflicting evidence. Patients produce fewer filled pauses than controls, as do their part-

ners when normalised by number of words. As filled pauses may indicate a wish to retain the floor during a turn (Clark & Fox Tree, 2002), it may be that patients are less likely to employ these turn holding techniques. A similar pattern of filled pauses in patients' partners demonstrates the impact of the presence of the patient on the behaviour of their interlocutors. The specific reason for this is unclear. It may be due to them aligning their own talk to the speech pattern of the patient (Garrod & Pickering, 2009) in a similar way to alignment in nonverbal behaviours (Lavelle, Howes, Healey, & McCabe, 2013). However, this possibility is contrary to evidence suggesting no differences in patterns of disfluencies between monologue and dialogue (Finlayson & Corley, 2012). It may also be indicative of the reduced competition for the floor in patient interactions, such that turn holding cues are less necessary.

The pattern of unfilled pauses supports this theory, with patients' partners producing more than either the patients they are interacting with, or controls, but not when normalised by number of words, suggesting this is a difference at the level of the turn. Taken together, this suggests a breakdown in turn-taking in dialogues containing a patient. Within-turn pauses may occur at points where patients' partners have reached a transition relevance place (TRP), where turn change is normally licensed (Sacks, Schegloff, & Jefferson, 1974), and are expecting (or encouraging) the patient to take the floor. In the control dialogues, turn-taking is undertaken smoothly, hence there are fewer unfilled pauses and more floor-holding filled pauses. Similarly, where a patient pauses, this may be taken as a turn-change cue, resulting in patients also producing fewer unfilled pauses. The effect would then only be apparent within patients' partners' turns, and could indicate that patients are less responsive to turn taking cues or more reluctant to select as next speaker. This explanation is consistent with the observation that patients produce fewer turns, evidence that patients are less able to coordinate their behaviour with others during interaction (Kupper, Ramseyer, Hoffmann, & Tschacher, 2015; Lavelle, Healey, & McCabe, 2014), and work that suggests that one of the social skills deficits in patients with schizophrenia manifests in poor turn-taking (Mueser et al., 1991). Note however that the 200ms cut off for unfilled pauses is arbitrary, and this analysis does not differentiate between long and short unfilled pauses which are expected to have different interactive consequences; for example short pauses may simply reflect within-turn phrasing, and not turn-taking issues per se.

In future work, we intend to exploit the fact that the current dataset gives us the opportunity to explore turn-taking in interactions between patients and partners who are unaware of their diagnosis directly and at a much finer-grained level. For example, we intend to examine unfilled pause distributions *between* speakers. Based on the preliminary results reported above, we would expect that there would be more 'inappropriate' turn changes in the dialogues with patients (characterised in opposition to the "no gap no overlap" model; Sacks

et al., 1974), in addition to the increase in within-turn pauses observed for patients' partners. Other turn-taking cues that may be less likely to be responded to by patients include nonverbal behaviours such as gesture and gaze, and future work will investigate these behaviours at points where there are unfilled pauses or potential TRPs.

The evidence suggests that smooth coordination of turns is problematic in patients' dialogues. We know that patients have difficulty coordinating their nonverbal behaviour with others, which is associated with difficulty building relationships (Kupper et al., 2015). Therefore patients' turn-taking difficulties may also contribute to their poor social functioning, which is one of the most debilitating and poorly understood aspects of schizophrenia. Understanding the nature of these deficits and their interactional relevance would provide a focus that could be targeted through psychosocial interventions, such as those that have proven effective in autism (Wert & Neisworth, 2003). It would also provide a measurable behavioural marker of social deficit, which could be monitored for improvement. This line of research would provide a step change in an area of great clinical need.

Conclusions

This unique data demonstrates that not only are there communication difficulties in schizophrenia but they also impact on social interactions more broadly, thus providing new insights into the social deficits of this complex disorder. The data also support the idea that disfluencies are communicative solutions, not problems.

Acknowledgments

References

Allwood, J., Nivre, J., & Ahlsén, E. (1990). Speech management: On the non-written life of speech. *Nordic Journal of Linguistics*, *13*(01), 3–48.

Anderson, A., Bader, M., Bard, E., Boyle, E., Doherty, G., Garrod, S., . . . Weinert, R. (1991). The HCRC map task data. *Language and Speech*, *34*(4), 351–366.

Bailey, K. G., & Ferreira, F. (2007). The processing of filled pause disfluencies in the visual world. In *Eye movements: A window on mind and brain* (pp. 485–500). Amsterdam, NLD: Elsevier Science & Technology.

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278.

Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., & Brennan, S. E. (2001). Disfluency rates in conversation: Effects of age, relationship, topic, role, and gender. *Language and Speech*, *44*(2), 123–147.

Brennan, S., & Schober, M. (2001). How listeners compensate for disfluencies in spontaneous speech. *Journal of Memory and Language*, *44*(2), 274–296.

Brugman, H., & Russel, A. (2004). Annotating multimedia/multi-modal resources with ELAN. In *LREC: The fourth international conference on language resources and evaluation* (p. 2065-8).

Clark, H. H., & Fox Tree, J. E. (2002). Using *uh* and *um* in spontaneous speaking. *Cognition*, *84*(1), 73–111.

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).

Doyen, S., Klein, O., Pichon, C.-L., & Cleeremans, A. (2012). Behavioral priming: It's all in the mind, but whose mind? *PloS One*, *7*(1), e29081.

Finlayson, I. R., & Corley, M. (2012). Disfluency in dialogue: An intentional signal from the speaker? *Psychonomic Bulletin & Review*, *19*(5), 921–928.

Garrod, S., & Pickering, M. J. (2009). Joint action, interactive alignment, and dialog. *Topics in Cognitive Science*, *1*(2), 292–304.

Ginzburg, J., Fernández, R., & Schlangen, D. (2014). Disfluencies as intra-utterance dialogue moves. *Semantics and Pragmatics*, *7*(9), 1–64.

Godfrey, J. J., Holliman, E., & McDaniel, J. (1992). SWITCHBOARD: Telephone speech corpus for research and development. In *Proceedings of IEEE ICASSP-92* (pp. 517–520). San Francisco, CA.

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.

Hough, J., & Purver, M. (2014). Strongly incremental repair detection. In *Proceedings of the 2014 EMNLP conference.* Doha, Qatar: ACL.

Howes, C., Hough, J., Purver, M., & McCabe, R. (2014). Helping, I mean assessing psychiatric communication: An application of incremental self-repair detection. In *Proceedings of the 18th SemDial (DialWatt)* (p. 80-89). Edinburgh.

Howes, C., Purver, M., Healey, P. G. T., Mills, G. J., & Gregoromichelaki, E. (2011). On incrementality in dialogue: Evidence from compound contributions. *Dialogue and Discourse*, *2*(1), 279–311.

Howes, C., Purver, M., McCabe, R., Healey, P. G. T., & Lavelle, M. (2012). Helping the medicine go down: Repair and adherence in patient-clinician dialogues. In *Proceedings of the 16th SemDial (SeineDial)* (pp. 155–156). Paris.

Johns, L. C., Rossell, S., Frith, C., Ahmad, F., Hemsley, D., Kuipers, E., & McGuire, P. (2001). Verbal self-monitoring and auditory verbal hallucinations in patients with schizophrenia. *Psychological Medicine*, *31*, 705–715.

Kay, S. R., Flszbein, A., & Opfer, L. A. (1987). The positive and negative syndrome scale (PANSS) for schizophrenia. *Schizophrenia Bulletin*, *13*(2), 261.

Kupper, Z., Ramseyer, F., Hoffmann, H., & Tschacher, W. (2015). Nonverbal synchrony in social interactions of patients with schizophrenia indicates socio-communicative deficits. *PloS One*, *10*(12), e0145882.

Lavelle, M., Healey, P. G., & McCabe, R. (2013). Is nonverbal communication disrupted in interactions involving patients with schizophrenia? *Schizophrenia Bulletin*, *39*(5), 1150–1158.

Lavelle, M., Healey, P. G., & McCabe, R. (2014). Participation during first social encounters in schizophrenia. *PloS One*, *9*(1), e77506.

Lavelle, M., Howes, C., Healey, P. G. T., & McCabe, R. (2013). Speech and hand movement coordination in schizophrenia. In *Proceedings of the TiGeR Tilberg gesture research meeting.* Tilburg.

Leudar, I., Thomas, P., & Johnston, M. (1992). Self-repair in dialogues of schizophrenics: Effects of hallucinations and negative symptoms. *Brain and Language*, *43*(3), 487 - 511.

Levelt, W. (1983). Monitoring and self-repair in speech. *Cognition*, *14*(1), 41–104.

Maclay, H., & Osgood, C. E. (1959). Hesitation phenomena in spontaneous english speech. *Word*, *15*(1), 19–44.

McCabe, R., Dooley, J., John, P., Healey, P. G. T., Cushing, A., Kingdon, D., . . . Priebe, S. (2016). Training to enhance psychiatrist communication with patients with psychosis (TEMPO): A cluster randomized controlled trial. *British Journal of Psychiatry.*.

McCabe, R., Healey, P. G. T., Priebe, S., Lavelle, M., Dodwell, D., Laugharne, R., . . . Bremner, S. (2013). Shared understanding in psychiatrist-patient communication: Association with treatment adherence in schizophrenia. *Patient Education and Counselling*.

Meteer, M., Taylor, A., MacIntyre, R., & Iyer, R. (1995). *Disfluency annotation stylebook for the switchboard corpus* (Tech. Rep.). Department of Computer and Information Science, University of Pennsylvania.

Mueser, K. T., Bellack, A. S., Douglas, M. S., & Morrison, R. L. (1991). Prevalence and stability of social skill deficits in schizophrenia. *Schizophrenia Research*, *5*(2), 167–176.

Nicholson, H., Eberhard, K. M., & Scheutz, M. (2010). "Um... I don't see any": The function of filled pauses and repairs. In *Proceedings of DiSS-LPSS* (pp. 89–92).

Sacks, H., Schegloff, E., & Jefferson, G. (1974). A simplest systematics for the organization of turn-taking for conversation. *Language*, *50*(4), 696–735.

Schegloff, E., Jefferson, G., & Sacks, H. (1977). The preference for self-correction in the organization of repair in conversation. *Language*, *53*(2), 361–382.

Wert, B. Y., & Neisworth, J. T. (2003). Effects of video self-modeling on spontaneous requesting in children with autism. *Journal of Positive Behavior Interventions*, *5*(1), 30–34.

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: Wiley.