
Article:
Walker, T. orcid.org/0000-0002-2583-7232, Christensen, H. orcid.org/0000-0003-3028-5062, Mirheidari, B. orcid.org/0000-0002-7797-2778 et al. (5 more authors) (2020) Developing an intelligent virtual agent to stratify people with cognitive complaints: A comparison of human-patient and intelligent virtual agent-patient interaction. Dementia, 19 (4). pp. 1173-1188. ISSN 1471-3012

https://doi.org/10.1177/1471301218795238

Developing an Intelligent Virtual Agent (IVA) to stratify people with cognitive complaints: a comparison of human-patient and IVA-patient interaction

Traci Walker[1], Heidi Christensen[2], Bahman Mirheidari2, Thomas Swainston[3], Casey Rutten3, Imke Mayer3, Daniel Blackburn3, and Markus Reuber[4]

Corresponding Author: Dr Traci Walker, Department of Human Communication Sciences, University of Sheffield, 362 Mushroom Lane, Sheffield S10 2TS
0114 222 2420

Abstract

Previous work on interactions in the memory clinic has shown that Conversation Analysis (CA) can be used to differentiate neurodegenerative dementia (ND) from functional memory disorder (FMD). Based on this work, a screening system was developed that uses a computerised 'talking head' (IVA) and a combination of automatic speech recognition and CA-informed programming. This system can reliably differentiate patients with FMD from those with ND by analysing the way they respond to questions from either a human doctor or the IVA. However, much of this computerised analysis has relied on simplistic, nonlinguistic phonetic features such as the length of pauses between talk by the two parties.

To gain confidence in automation of the stratification procedure, this paper

[1] Human Communication Sciences, University of Sheffield
[2] Department of Computer Science, University of Sheffield
[3] Sheffield Institute for Translational Neuroscience (SITraN), University of Sheffield
[4] Academic Neurology Unit, University of Sheffield, Royal Hallamshire Hospital, Sheffield, UK

investigates whether the patients' responses to questions asked by the IVA are qualitatively similar to those given in response to a doctor. All the participants in this study have a clear FMD or ND diagnosis.

Analyses of patients' responses to the IVA showed similar, diagnostically-relevant sequential features to those found in responses to doctors' questions. However, since the IVA's questions are invariant, its use results in more consistent responses across people - regardless of diagnosis - which facilitates automatic speech recognition and makes it easier for a machine to learn patterns. Our analysis also shows why doctors do not always ask the same question in the exact same way to different patients. This sensitivity and adaptation to nuances of conversation may be interactionally helpful; for instance, altering a question may make it easier for patients to understand. While we demonstrate that some of what is said in such interactions is bound to be constructed collaboratively between doctor and patient, doctors could consider ensuring that certain, particularly important and/or relevant questions are asked in as invariant a form as possible to be better able to identify diagnostically-relevant differences in patients' responses.

1 Background to the study

The early recognition of neurodegenerative disorders is important to give clinicians an opportunity to treat patients before irreversible changes have occurred in the brain, but represents a considerable clinical challenge. Investigations such as Amyloid PET scans or tests on the cerebrospinal fluid are capable of detecting abnormalities when clinical manifestations are subtle or have not been noticed, but are expensive, and not suitable for screening purposes. Reuber et al. (2018) showed that it was possible to discern profiles based on a close study of patients conversational behavior, which could support the diagnostic differentiation between progressive neurodegenerative dementias (ND) and (non-progressive) functional memory disorders (FMD).[1] These profiles, created using conversation analysis (CA) techniques, are based on important

differences in the sequential and grammatical details of the talk patients produce in response to doctors' questions during memory clinic interactions.[2] For instance, patients can be differentiated by whether they are able to respond to compound questions, or the level of detail they give when they tell the doctor about their experiences of memory failures. With a view to streamlining the care pathway for people with memory complaints in the UK, we began a programme of research designed to automate this work. Mirheidari, Blackburn, Reuber, Walker, and Christensen (2016) showed that it is possible to extract automatically certain features of interaction initially detected using ''manual'' conversation analysis. Using manually-produced transcripts and machine learning techniques, a computer can predict independently-formulated best medical diagnoses with an accuracy of 90%.

The latest step has seen the introduction of an Intelligent Virtual Agent (IVA) to interact with patients instead of a human doctor (Mirheidari et al., 2017a); this work has shown that patients can be reliably differentiated based on a computerised analysis of their interaction with the IVA. Participants were recruited from among patients attending neuropsychological testing procedures for memory problems in the memory or neuropsychology clinics in a large city hospital in the UK[5]. They were given verbal and written instructions about how to talk to the IVA, which is a computerised talking head on a laptop screen. The IVA asked the participants a set of 12 recorded (ie., not computer voice synthesised) questions, developed based on the findings of Jones et al. (2015) and Elsey et al. (2015). The participants' responses were both audio- and video-recorded. So far, the IVA does not employ synthesised speech, nor does it prompt the user for responses or have any interactive features

[5] Data collection was approved by the London - City and East research ethics committee, REC reference 16/LO/0737.

(although its eyes blink, and its mouth is synchronised with the recorded questions). It simply mouths a previously recorded question which participants were encouraged to answer by speaking to the computer. The user controls when the IVA speaks by pressing a button to go on to the next question.

Our ongoing work has the joint aims of reducing the ASR error rate, improving and customising the IVA interface, and expanding the number of diagnostic categories the system can reliably distinguish -- specifically to include patients with mild cognitive impairment (MCI), i.e., patients likely to develop dementia (see Mirheidari, Blackburn, Walker, Reuber, & Christensen, under review). The ultimate goal of this research is to reduce the burden on memory clinics throughout the UK by developing an automatic stratification process which allocates people presenting with memory complaints into the most appropriate care pathway more quickly, and reduces the need for unwarranted invasive tests.

To date our work has shown convincingly that we can automatically differentiate patients with FMD from those with ND by analysing the way they respond to questions from either a human doctor or the IVA. However, much of the computerised analysis of patients' responses to the IVA has relied on simplistic, nonlinguistic phonetic features such as the length of pauses between talk by the two parties. In this paper we explore to what extent patients' responses to the IVA qualitatively replicate responses to questions from (human) doctors, and how qualitative observations about patients' conversational contributions can lead to further improvements of the automated differential diagnostic process. To gain confidence in automation of the stratification procedure, we need to know whether

patients' responses to questions posed by a computerised talking head -- the IVA -- are similar in kind to those given to human co-participants, on which the programming of the automated analysis was based. To find this out, however, we start with an investigation of the questions asked by doctors in diagnostic memory clinic encounters. Our analysis of these questions reveals that there is considerable variability in the linguistic forms doctors use to ask one of the most important questions: about patients' most recent memory failure[6]. We show how the conversation prior to this particular enquiry constrains the question formats available to the doctors, and also demonstrate the subtle but important effect the different question formats have on the form of the patients' responses; eg., how patients (both with FMD and with ND) match their responses to the particular way in which a question is asked.

We then turn to the IVA interactions, and show that here too, patients design their responses to fit the IVA's question. These answers, however, show striking linguistic similarities regardless of diagnosis; we argue that this is due to the invariance of the IVA's question. Despite this greater similarity in patient response, we still pick up the qualitative diagnostic pointers found in the doctor-patient interactions (Mirheidari et al., 2017a, 2017b, under review), and can successfully classify patients' responses as typical of FMD or ND. The paper concludes with a discussion of some of the advantages IVA interactions provide.

2 Methodology

The patient recordings used in this study come from two different corpora, both

[6] Notably this variability did not preclude the successful differentiation of the patients' responses into those typical of the FMD and ND groups.

recorded in memory clinics in the same UK city hospital. The data collection process for interactions with the IVA was described above, and more detail can be found in Mirheidari et al. (2017). The data of interactions between human doctors and patients was collected for an earlier study and is described in more detail in Jones et al. (2015) as well as Reuber et al. (2018). In this paper we analyse the interactions of a purposively selected subset of 22 patients drawn from each of these datasets for whom we have neuropsychological data and independently generated, clinically firm medical diagnoses of FMD or ND. We compare 10 interactions between patients and the IVA (5 with FMD and 5 with ND) and 12 interactions of different patients (6 with FMD and 6 with ND) with a neurologist (either a fully trained consultant or senior trainee -- ‘‘Speciality Registrar’’). Participant demographics are shown in tables 1-2.

Table 1: Neuropsychological data for patients interacting with the IVA

	FMD	ND
Age	55.5 ($\pm$ 3.99)	63.8 ($\pm$ 4.55)
M:F ratio	3:2	3:2
ACE-R	84.17 ($\pm$ 6.26)	58.2 ($\pm$ 11.85)

Table 2: Neuropsychological data for patients interacting with consultants

	FMD	ND
Age	58.83 ($\pm$ 4.78)	64.0 ($\pm$ 4.39)
M:F ratio	3:3	1:5
ACE-R	92.83 ($\pm$ 1.52)	73.83 ($\pm$ 4.72)

We focus here on responses to a history-taking question which proved particularly important in the ‘‘manual’’, qualitative differentiation of FMD and ND interactions: asking the patients to recall a recent memory failure. Elsey et al. (2015, pp 1074ff) showed that FMD patients consistently provide relevant and detailed examples, whereas patients with ND struggle to respond, and offer vague, generic responses

such as ‘‘all the time’’.

Therefore, this paper employs the methodology of conversation analysis (CA) to investigate whether the patients' responses to questions asked by the IVA are qualitatively similar to those given in response to a doctor. For a conversation analyst (as well as for the participants), the central question about what is happening in talk is *why that now?* CA aims to show how a turn at talk exhibits design features that show -- or markedly do not show -- an orientation to what was said before in terms of the action done by the turn (the why), the words used in the turn (the that), and the sequential placement of the turn (the now). The sequence of turns in any given interaction is of primary concern in CA. Analysis proceeds by examining how a particular turn at talk displays an understanding of the action being done by the preceding turn (Levinson, 2013; Schegloff, 2007). Any given turn is not only shaped by the prior talk, but also contributes to the sequence in which the next turn will be understood (Heritage, 1984; Sidnell, 2010). In other words, the analysis always considers how context shapes the talk; but it also considers how the talk shapes the context to be just as important.

CA has consistently proven to be a useful technique in investigating the crucial role of communication in medical history-taking and medicine prescribing (eg., Heritage & Maynard, 2006, Stivers, 2005a, 2005b), and has shown the importance of analysing the fine details of talk -- that is, the actual words used, the grammatical forms, the length of pauses, the talk preceding and the talk following. Indeed, Heritage, Robinson, Elliott, Beckett, and Wilkes (2007) show the effects of changing just one word in a medical history taking encounter, and Reuber, Monzoni, Sharrack, and Plug (2009) describe how linguists using CA could successfully differentiate between

patients with a diagnosis of epilepsy or psychogenic non-epileptic seizures, based only on a modified history-taking protocol. This paper, and the larger project which it comes from, continues in that vein.

3 Analysis

As noted in previous research (Elsey et al., 2015, 1074), doctors' questions are produced with a variety of forms: ‘‘. . . the neurologist asked the patient . . . ‘can you give me an example of the last time your memory let you down?', or some variant thereof'' [emphasis added]. Even though different linguistic forms were used to prompt patients to describe their memory problems, the original analyses show that the differences in the responses relate to the participants' diagnoses -- not to the way the question was asked. In the data sets we analyse here, the form of the IVA's request to tell about a recent memory failure is (of course) invariant: it always asks ‘‘Tell me what problems have you had with your memory.'' This too yielded robust results -- the answers to this question as asked by the IVA contribute heavily to differentiating FMD from ND.

Following CA methodology, we begin our analysis of the patients' responses with a consideration of the question that preceded them. First we describe the structural differences of the question formats employed by the doctors. We then show how an analysis of the sequential location of the doctors' questions offers a potential explanation for the variety of linguistic forms employed to accomplish the activity of asking the patients to describe a recent memory failure. Patients' responses to the doctor (regardless of their diagnoses) exhibit an orientation to the subtle differences in the doctors' questions. The IVA's question cannot be affected by the sequence-so-far; it is not programmed to select a next utterance based on what the patient has

previously said. We show how this invariance results in an interesting similarity in almost all the responses, again regardless of diagnosis: the use of the words ‘forget’ or ‘remember.’

3.1 Doctors’ question formats

In our subset of twelve interactions between doctors and patients, the only format that doctors use more than once is ‘‘can you tell me the last time your memory let you down’’. Even this is only produced two times. A variety of other grammatical formats are employed to ask the patients about recent memory failures. Sometimes the noun example is employed, with a difference in the number requested: ‘‘an example’’ vs ‘‘examples’’. The potential difficulty of responding to the question is acknowledged once by including a quantifier,‘‘can you give me any examples’’; there is also a case in which the doctor uses an adjective to specify the type that should be provided, ‘‘can you give me any significant examples’’. Different forms of modal verbs are used: ‘‘can you tell . . . ’’ vs ‘‘could you tell . . . ’’, and once the word ‘‘failed’’ is employed in place of ‘‘let you down’’. These form a group of grammatical alternatives; eg., plural vs singular, modified vs unmodified nouns.[3] In some cases, however, radically different formats are used to accomplish the activity of requesting information about the patients’ experience of memory problems, eg., ‘‘what kind of things do you find that you’re struggling with.’’ Additionally, in some cases the doctors do not appear to ask for such information. In the following sections we briefly consider each of these different options of enquiry in turn.

3.1.1 Grammatical alternatives

The following examples show variations in the use of singular vs plural nouns as well as different modal verb forms. Please see the appendix for transcription conventions.

(1) 056: DR-ND

```
1 →  Neu:    okay ((pause)) and- (.) could you give me an example of the
2            last time your memory let you down
3            (0.3)
4     Pat:   no (1.0) no (1.0) can you ((turns to partner))
```

(2) 036: DR-FMD

```
1 →  Neu:    um (1.4) when: (0.4) can you give me an example of the last
2            time your memory let you sa- let you down
3            (1.0)
4     Pat:   .hhhhh i- m- erm:: (0.5) I was on the telephone (0.8) making an
5            appointment for someone at wor:k .hh and we send appointment
6            cards out to 'em .hh and I picked the appointment card up and I
7            said to the- (.) customer .hh right I will send you a and I
8            couldn't think (0.9) of the word appointment ca(h)rd
9            ((laughs))=
```

(3) 083: DR-ND

```
1 →  Neu:    and cn- can you tell: m:e=give me examples of how your memory
2            has let you down
3            (2.3)
4     Pat:   I find it- I find it very easy to forget things these days
```

Despite the (admittedly subtle) different grammatical formats -- ‘‘could’’ and ‘‘can’’ with the singular ‘‘example’’; ‘‘can’’ with the plural ‘‘examples’’ -- all these questions perform a similar action. As shown by Elsey et al. (2015), the reponses are one way of differentiating ND from FMD: here, the ND patients (in examples 1 and 3) display an inability to respond adequately, whilst the FMD patient (example 2) provides a detailed example, as requested.

3.1.2 Radically different formats

Sometimes the doctors’ probes for a description of the patients’ memory problems are radically different from ‘‘can you tell me the last time your memory let you down’’ or any of the examples provided above, as shown by the arrowed turns below.

(4) 043: DR-ND

```
 1 →  Neu:    okay. An:d are there any kind of specific things that you can
 2 →          think of that- th'kind of (.) typical memory problems that you
 3 →          were having at that time
 4            (0.4)
 5    Oth:    I don't know (now)=
 6    Pat:    =I can't remember ((la[ughs))
 7    Neu:                        [n:o problem ha that's OK don't worry and
 8 →          what about more recently<what kind of things do you find that
 9 →          you're struggling with
10            (3.1)
11    Pat:    er
12            (2.6)
13    Oth:    well I mean she'll ask me for (0.9) er (0.7) what I want for me
14            tea and then ask same question
15    Neu:    mm hm
16            (0.7)
17    Oth:    what (0.4) two minutes later or something like that
```

(5) 034: DR-FMD

```
 1 →  Neu:    okay so::- I have read through your GP's letter but would you
 2 →          like to describe the problem now: with which you are seeing him
 3            today
 4            (0.3)
 5    Pat:    tch .hhhhhh well (0.8) I- I seem to have (0.5) more long-term
 6            (.) memory loss (0.8) um: (0.3) more than anything else really=
 7    Neu:    =yeah
 8    Pat:    Short-term is spasmodic in't it
 9            (1.8)
10    Pat:    I- I have (0.3) >I mean< I wo:rk for myself I have guys
11            wo:rking for me (0.6) and it's quite frustrating because (0.5)
12            I-I'll sa:y (0.4) >obviously said< something to them at w- at
13            some stage (0.8) and then when I go out and check what they're
14            doing I'll say why are you doing that
```

These arrowed turns can be seen to be doing the same action as the arrowed turns in examples 1, 2 and 3. There are aspects of the design of the turns that prompt details of any memory difficulties from the patients, namely ‘‘the kind of typical memory problems that you were having’’; ‘‘would you like to describe the problem’’. Additionally, and crucially from the conversation analytic standpoint, the participants respond to these turns as requests to tell about their memory problems. Due to their different underlying problems, however, they give different responses. In example 4, the patient with ND states that she ‘‘can't remember’’ (line 6), and her partner responds to the follow up question with a specific example. In example 5, the patient with FMD initially gives a response which specifies the type of memory loss he is

experiencing, but goes on to provide an example (lines 10-14).

3.2 Doctors' use of -- and patients' orientation to -- differences in question formats

In this section, we show how it is that the doctors come to use such a variety of question formats. We also show that patients design their responses to 'match' the format of the doctors' questions -- whether these are grammatical alternatives or more radical differences.

Minute linguistic calibrations to the format of the doctors' questions can be seen in the previously presented examples 1 and 3. In (1), the patient responds to ''could you give me an example . . . '' with ''no.'' This response orients to the fact that ''could you'' may be used to introduce yes/no (polar) questions -- not just to request that an example be given. It is therefore fitted not only to the linguistic sequence in which it is placed (responding to what is formally a polar question), but also demonstrates her inability to participate in the action sequence. By producing only the bare "no", she shows that she is unable to collaborate in the requested activity.

In (3), the doctor asks for ''examples'' -- plural. Although the patient's response is in this case not well-fitted to other aspects of the question format, he reports that he finds it ''very easy to forget things'', which could display an orientation to providing more than one example in his response. Though we are not arguing that this conforms to the question by actually providing examples, we do wish to note the use of plural nouns in both the question and the response.

The following examples provide other evidence of the way that the sequence affects the form of the doctors' enquiries. In example 6, the patient describes her memory problems in response to the doctor asking about ''difficulties.''

(6) 100: DR-ND

```
 1 →  Neu:   okay (0.3) and- (0.3) tell me a little bit about (0.3) the
 2 →         difficulties that you're experiencing and your symptoms
 3           (2.0)
 4    Pat:   well (0.6) just forgetting is er (0.4) it's just er (2.0)
 5           difficult to (0.3) manage my day (1.8) and erm (1.1) so then I
 6           went the GP °to see what- °what's what's happening because of
 7           the (1.6) forgetting=
 8    Neu:   =mm hm
 9    Pat:   and (.) so many (0.3) times go upst- (upin er) er (0.4) er
10           (0.5) to go upstairs and then downstairs because of (0.2) when
11           (0.6) I needed to pick it up something and then the (0.7) and
12           then (1.7) whenever I get to to t'the place that [I think that
13    Neu:                                                    [yeah
14    Pat:   they're (0.3) so I forgotted why I ca: came to (0.4) to do that
15           er
             1m 48s omitted
16    Neu:   can you give me: any- (0.2) significant examples where your
17           memory's let you down recently=anything that may have been=been
18           upsetting or embarrassing anything that stands out
19           (1.2)
20    Pat:   oh that's hard ((laughs - turn to companion)) you know (what
21           that) more than .hh but er there is er (0.4) um: embarrassing
22           (2.5) try to (0.6) er express what I want to say and then (0.8)
23           and then the words just goes (0.8) a[way and then er (0.3) it's
24    Neu:                                       [okay
25    Pat:   just like (oh wait a) minute I know what I want to say so but
26           then sa- (0.4) it's gone
27    Neu:   °okay°
28    Pat:   that- that kind of situation like that
```

In lines 9-15, the patient explains that when going upstairs or downstairs she often forgets what she went there to get. After some talk designed to establish how long the patient has been noticing memory problems (data not shown), the doctor moves to the activity of securing some examples. To do this, in this sequential location, she uses an adjective to modify ''examples'', and extends her turn with another clause: ''can you give me any significant examples where your memory's let you down recently anything that may have been upsetting or embarrassing, anything that stands out''. By adding these characteristics, the doctor displays that she has been listening to, and has understood the patient's prior talk as giving (some) examples. By using a modifier

and a turn extension she alters her request to encompass other examples that would meet these additional criteria.

The patient in example 7 also provides a lengthy (2.5 minutes) and detailed description of his memory problems. This occurs in response to the second stage of the interaction in which the doctor asks why the patient has come to the memory clinic. For reasons of space, only part of this response is reproduced below.

(7) 011: DR-FMD

```
 1    Neu:    d'you'wanna tell me:: (1.0) um why you've come today and what
 2            expectations you have about the clinic
 3    Pat:    .hh well one of the reasons was because I have a partner (0.7)
 4            a:nd he was sort of reminiscing abou:t (0.2) times past>like
 5            holidays and things we've had and I thought .h well I can't
 6            remember tha:t an' I can't remember that happening .hh An'
 7            the's there's other things where, cos I work in a public
 8            hou:se: (0.8) I'd be down stairs working (0.3) .h an' then
 9            somebody'll say oh a- a pri:me example was er on Frida:y (.)
10            .hh ((clears throat)) when I needed to go upstairs for
11            something (1.0) tkh an' I just set off going and we've got a
12            telephone of the staircase going upstairs .h an' just as I went
13            upstairs the phone rang (.) .hh I: had to do- somebody said oh
14            would you mind doing a quick survey=so I did this quick survey
15            .hh and I went to the top of the stairs (0.2) and thought (1.3)
16            what have I come up here for
```

Here, the patient provides examples of going up or down stairs and forgetting what he went to get, and of forgetting a task -- after being interrupted to do a telephone survey. He continues (data not shown) with clear descriptions of having to write down invoice amounts in order not to forget them whilst counting money (see also the extended discussion of this extract in Jones et al., 2015). He then proceeds to describe his expectations of the visit, after which he gives yet another example of his perceived memory problem. He marks the return to this topic by prefacing his turn with ''another thing as well'', going on to detail how he has forgotten to bring his prescription medications with him, as instructed on his appointment letter. In this interaction, the doctor does not attempt to ask for any additional examples, and it is difficult to see how he could whilst still presenting as a competent conversational partner. Of his own accord, the patient provides so many examples, in so much detail,

that to ask for others could be taken as evidence by the patient that the doctor was not actually attending to what he was saying.

3.3 Responses to the IVA

In analysing the subset of 10 interactions of patients and the IVA, we found that both patients diagnosed with ND or FMD responded to the IVA's questions in a similar way; however this does not obscure the diagnostically relevant differences in their responses. Namely, all but two of the responses to ''Tell me, what problems have you had with your memory'' use some form of the word 'forget' or 'remember.' In the following examples, we have included (mindful of space considerations) all of the talk the patients' produced in response to the IVA's question (that is, everything that was said up to questions by the participants about to how to move on to the next question).

(8) P01: IVA-FMD

1	Pat:	er (.) my memory er sh:ort term memory (0.5) when people are
2		talking to me .hhh I'm **forgetting** (.) er (0.3) within (.) seconds
3		what they said=or the conversation .hh I can't take things in
4		(0.2) .hhhh er long term memory's not too bad but sh:ort term

(9) P03: IVA-FMD

1	Pat:	ah various problem:s (1.2) with my memory .hhhhhh erm: m-s- a
2		lot of the ti:me I don't **remember** (0.3) what- .hhhhh (0.3) I've
3		got up to do: hhhhh erm .hhhhh go shomewhere:: (0.6) get there:
4		don't know what I've gone for (1.7) erm (1.0 can't **remember**
5		wor:d:s (1.0) can usually **remember** the er (0.5) the letter they
6		start with .hhhhhh (0.4) but not the words
		1 minute 19 seconds omitted, giving additional examples
7		but yeah it happens a lot

(10) P13: IVA-ND

1	Pat:	u:m (2.6) a loss of um (2.8) **remembering** things at times (0.4)
2		words (.) different places (10.2) words and different places
3		just re- just remembering (1.5) remembering things at time
4		(0.4) not easy

(11) P11: IVA-ND

1	Pat:	I **forget** everything

The patients in examples 9 and 10 both use the term 'remember', whilst in 8 and 11

they use ‘forget.’ However, the patients in 10 and 11 have a diagnosis of ND, whilst the patients in examples 8 and 9 do not. In fact, all the participants diagnosed with ND use a form of these words in their responses. Only one response did not include one of these terms, and comes from a participant with a diagnosis of FMD. His response is ‘‘erm (0.4) I lose words’’, which serves to describe a specific problem that he has with his memory -- as asked for by the question.

The following example shows a participant responding to the IVA’s question by asserting that his problems were due to his confidence, not memory. In other words, he is denying the presupposition of the question. Despite this, he also uses the word ‘forgotten’.

(12) P17: IVA-FMD

```
1    Pat:    my problems weren't actually to do with memory it was my
2            confidence (.) an::d I was making assumptions: when I was right
3            (0.4) that in fact (.) I had made mistakes I had forgotten to do
4            something (0.5) er:: so: I was misidentifying it as memory
```

In example 12 the patient is explaining that his confidence issues had led him to believe, incorrectly, that he had forgotten something (‘‘I was making assumptions . . . [that] I had forgotten’’). The other participants, with both ND and FMD, all use the words ‘forget’ or ‘remember’ to describe ongoing, unresolved issues. It is possible that even though this participant’s memory issues are resolved, the sequential constraints of the question make it relevant for him to use one of the same words. The words ‘forget’ and ‘remember’ clearly link to the term ‘‘memory’’ in the IVA’s question. The fact that these words are used by participants regardless of their diagnosis shows that, as in the doctor-patient data, all the participants orient to maintaining coherence between their talk and the question that was put to them. Similarity across the responses is not too surprising from a conversation analytic standpoint, since the IVA’s question is the same every time. One thing the similarity

of the responses shows, however, is that the IVA is being treated in a similar way to a human co-participant despite the fact that it displays no cohesion between its questions, nor any orientation to the responses.

It is also clear, however, that despite the similarities in the responses, there is a detectable difference in the way that participants are able to respond to the action enacted by this question[5]. What is more, we can observe the diagnostically relevant differences described by (Elsey et al., 2015, 1074) in examples 8-11.The patients with ND (examples 10 and 11) offer a ‘‘routine or common problem, rather than a specific incident’’, not expanding their responses much beyond phrases containing the words ‘forget’ or ‘remember,’ whilst those with FMD (examples 8 and 9) use ‘forget’ or ‘remember’ in prefaces to extended responses that give ‘‘a relevant and detailed example of a particular recent event.’’

4 Discussion

Starting with the knowledge that patients with ND and FMD can be reliably differentiated by their responses to questions asked during the history-taking phase of memory clinic interactions, this paper focusses on the formats employed to ask one particular, diagnostically-relevant question. We then analysed the relationship between the responses to these subtly different question formats, and compared them to patients’ responses to an invariant questions asked by an IVA presented on a computer screen. Our analyses of the patients’ responses to the IVA showed similar sequential features to those found in responses to doctors’ questions. Although the answers given by patients with ND and those with FMD were more similar in response to the IVA’s questions, which were uninformed by previous talk in the interaction, the differences in responses between patients with ND and FMD are still

reliably present.

Human analysts can work through individual differences to find systematic patterns, as shown in the initial work by Elsey et al. (2015); Jones et al. (2015), and best in Reuber et al. (2018). However, this work was carried out by trained conversation analysts and does require time and intensive training. Our ongoing research programme indicates that a computer-based approach can produce similar results, but the programming (including how to put particular questions) could benefit from better input data. This will be of particular importance for the reliable differentiation of the very subtle differences in interactional performance expected between patients with FMD and those with the mildest, earliest manifestations of neurodegenerative cognitive impairment (i.e. patients with MCI).

The use of an IVA demonstrably creates an environment where the responses are more consistent across people, which will facilitate automatic speech recognition and make it easier for a machine to learn patterns. The nearly systematic use of the words 'remember' and 'forget' in response to the IVA will also aid automatic speech recognition (ASR) and machine learning algorithms by providing multiple examples of the same word, spoken both by people with and without neurodegenerative disease. In the dataset of interactions with human doctors, there were large variations in terms of length of interaction and how much the doctors spoke, as well as how they designed their questions, making it more difficult for a machine even to identify the doctors question automatically, never mind interpret the patients answers.

For diagnostic or stratification purposes, it would be best if patients' responses were

not unduly affected by the preceding talk. For instance, in example 7 there is no need for the doctor to ask ‘‘can you give me an example of the last time your memory let you down’’ since the patient has already provided one. Nevertheless, one could also argue that, for rigour and replicability, we must compare like with like -- the same questions with the same responses. Following this line of reasoning, the doctor should ask the question, regardless of how sequentially ill-fitted it is. However, this paper has shown that it would be unrealistic for doctors to ignore the sequence-so-far.

It is important to point out that doctors’ sensitivity and adaptation to nuances of conversation is not, in all contexts, a bad thing. For instance, slight alterations of a particular question may make it easier for patients to understand. Mirheidari et al. (under review) provides some indication that human doctors pick up some of the interactional differences between patients with ND or FMD in that they speak differently to patients from these two groups (as measured by the number of unique words they use) even before they have formally made a diagnosis. However, doctors may modify their conversational style without becoming consciously aware of it, or without recognizing the change in their own conversation profile as a diagnostic indicator. It is unclear whether the doctors’ adaptation of their questions to the conversational context ultimately helps or hinders diagnostic categorisation, but it points to a need to recognise, and for diagnostic purposes filter out, at least some of the collaborative contribution of the doctors’ talk.

The IVA does not face the same challenges. Its questions are unaffected by conversational constraints resulting from preceding talk. It therefore provides a potential solution to this interactional bind. The IVA’s questions are invariant, and the

answers to its questions therefore provide data that may be more suitable for comparative diagnostic purposes. Indeed, like the clearly separate steps of a standardized physical examination, each question posed by the IVA represents a new challenge, whereas the emergent form of doctors questions in human-to-human interactions is the result of a complex conversational collaboration with the patient. Although the main intention of the present research was to contribute to the development of an effective automatised stratification tool for people with cognitive impairment, our findings may also inform the communication strategy of doctors seeing patients in conventional human-to-human clinic settings. While we demonstrate that some of what is said in such interactions is bound to be constructed collaboratively between doctor and patient, doctors could consider ensuring that certain, particularly important and/or relevant questions are asked in as invariant a form as possible to be better able to identify diagnostically-relevant differences in patients' responses. Since humans are (at least currently!) still better than computers at picking up and analysing a wider range of nuances in talk-in-interaction, and better able to cope with any variability than an automated system, the use of invariant (or nearly so) question formats should add to their ability to interpret the response, compare it to others, and extract the most diagnostically relevant information.

5 Strengths and limitations

This study shows that patient responses to questions posed by a computerised talking head share important sequential features with responses to a similar question as posed by human doctors. Since the computer's questions are invariant, the answers to its questions provide data that may be more suitable for comparative diagnostic purposes. Only a small subset of data was analysed, focussing on only one question. This number of cases is adequate for the qualitative method employed (Conversation

Analysis), but CA could also be employed on a larger dataset, looking at responses to a variety of questions. In addition, we do not have PET or CSF results for any of the participants (although our medical diagnoses are based on current routine neurology practice and we have not included any patients whose diagnoses were in any way uncertain).

6 Conclusions

In conclusion, this paper shows that participants interacting with an IVA treat it like a normal co-participant in designing their talk -- in fact one even chastises it for not orienting to what it's been told before. Using an IVA to 'chat' with patients about their memory concerns provides greater control over question format, independent of prvious talk, and thus ensures comparability of the responses; a necessary step in refining and fully automating the diagnostic differentiation procedure.

Notes

[1]Functional memory disorder is a benign albeit distressing condition with no underlying 'organic' cause (Schmidtke, Susanne, & Metternich, 2008). People suffering from FMD experience real memory problems in their everyday lives, but they are generally not found to be suffering from depression (which is a well-known cause of memory problems) and typically score within normal limits on neuropsychological tests.

[2]Recent work has also incorporated automatic analysis of lexical and acoustic features, see Mirheidari et al. (under review)

[3]We include ''failed'' vs ''let you down'' here due to their semantic equivalence.

[4]Inspection of the video files shows the doctor looking in the direction of the patient as he takes this turn at talk.

[5]Mirheidari et al. (2017a, 2017b) make use of features extracted from answers to this question in successfully using an automatic analysis to differentiate the patient groups.

References

Elsey, C., Drew, P., Jones, D., Blackburn, D., Wakefield, S., Harkness, K., . . . Reuber, M. (2015). Towards diagnostic conversational profiles of patients presenting with dementia or functional memory disorders to memory clinics. *Patient Education and Counselling,* 98, 1071-1077.

Heritage, J. (1984). *Garfinkel and ethnomethodology*. Cambridge: Polity Press.

Heritage, J., & Maynard, D. W. (Eds.). (2006). *Communication in medical care.* Cambridge: Cambridge University Press.

Heritage, J., Robinson, J. D., Elliott, M. N., Beckett, M., & Wilkes, M. (2007). Reducing patients' unmet concerns in primary care: the difference one word can make. *Journal of General Internal Medicine*, 22(10) 1429-1433.

Jones, D., Drew, P., Elsey, C., Blackburn, D., Wakefield, S., Harkness, K., & Reuber, M. (2015). Conversational assessment in memory clinic encounters: interactional profiling for differentiating dementia from functional memory disorders. *Aging and Mental Health*, 20(5), 500-509.

Levinson, S. C. (2013). Action formation and ascription. In J. Sidnell & T. Stivers (Eds.), *The Handbook of Conversation Analysis* (p. 103-131). Oxford: Wiley-Blackwell.

Mirheidari, B., Blackburn, D., Harkness, K., Walker, T., Venneri, A., Reuber, M., & Christensen, H. (2017a). An avatar-based system for identifying individuals likely to develop dementia. In Proceedings of Interspeech 2017 (p. 3147-3151).

Mirheidari, B., Blackburn, D., Harkness, K., Walker, T., Venneri, A., Reuber, M., & Christensen, H. (2017b). Toward the automation of diagnostic Conversation Analysis in patients with memory complaints. *Journal of Alzheimer's Disease*, 58, 373-387.

Mirheidari, B., Blackburn, D., Reuber, M., Walker, T., & Christensen, H. (2016).

Diagnosing people with dementia using automatic Conversation Analysis. In Proceedings of Interspeech-2016.

Mirheidari, B., Blackburn, D., Walker, T., Reuber, M., & Christensen, H. (under review). Dementia detection using automatic analysis of conversations. *Computer Speech and Language*.

Reuber, M., Blackburn, D., Elsey, C., Wakefield, S., Ardern, K. A., Harkness, K., . . . Drew, P. (2018). An interactional profile to assist the differential diagnosis of neurodegenerative and functional memory disorders. *Alzheimer Disease and Associated Disorders.* doi: 10.1097/WAD.0000000000000231

Reuber, M., Monzoni, C., Sharrack, B., & Plug, L. (2009). Using interactional and linguistic analysis to distinguish between epileptic and psychogenic non-epileptic seizures: a prospective, blinded multirater study. *Epilepsy and Behavior*, 16, 139-144.

Schegloff, E. A. (2007). *Sequence Organization in Interaction: A Primer in Conversation Analysis Volume 1*. Cambridge: Cambridge University Press.

Schmidtke, K., Susanne, P., & Metternich, B. (2008). The syndrome of functional memory disorder: definition, etiology, and natural course. *The American Journal of Geriatric Psychiatry*, 16(12), 981-988.

Sidnell, J. (2010). *Conversation Analysis: An Introduction*. Oxford: Wiley-Blackwell.

Stivers, T. (2005a). Non-antibiotic treatment recommendations; delivery formats and implications for patient resistance. *Social Science and Medicine*, 60, 949-964.

Stivers, T. (2005b). Parent resistance to physicians' treatment recommendations: one resource for initiating a negotiation of the treatment decision. *Health Communication*, 18(1), 41-74.