

Steering the Conversation: A Linguistic Exploration of Natural Language Interactions with a Digital Assistant during Simulated Driving

David R Large¹, Leigh Clark¹, Annie Quandt¹, Gary Burnett¹, Lee Skrypchuk²

¹Human Factors Research Group, Faculty of Engineering, University of Nottingham, UK

²Jaguar Land Rover Research, International Digital Laboratory, Coventry, UK

Abstract

Given the proliferation of ‘intelligent’ and ‘socially-aware’ digital assistants embodying everyday mobile technology – and the undeniable logic that utilising voice-activated controls and interfaces in cars reduces the visual and manual distraction of interacting with in-vehicle devices – it appears inevitable that next generation vehicles will be embodied by digital assistants and utilise spoken language as a method of interaction. From a design perspective, defining the language and interaction style that a digital driving assistant should adopt is contingent on the role that they play within the social fabric and context in which they are situated. We therefore conducted a qualitative, Wizard-of-Oz study to explore how drivers might interact linguistically with a natural language digital driving assistant. Twenty-five participants drove for 10 minutes in a medium-fidelity driving simulator while interacting with a state-of-the-art, high-functioning, conversational digital driving assistant. All exchanges were transcribed and analysed using recognised linguistic techniques, such as discourse and conversation analysis, normally reserved for interpersonal investigation. Language usage patterns demonstrate that interactions with the digital assistant were fundamentally social in nature, with participants affording the assistant equal social status and high-level cognitive processing capability. For example, participants were polite, actively controlled turn-taking during the conversation, and used back-channelling, fillers and hesitation, as they might in human communication. Furthermore, participants expected the digital assistant to understand and process complex requests mitigated with hedging words and expressions, and peppered with vague language and deictic references requiring shared contextual information and mutual understanding. Findings are presented in six themes which emerged during the analysis – *formulating responses; turn-taking; back-channelling, fillers and hesitation; vague language; mitigating requests* and *politeness and praise*. The results can be used to inform the design of future in-vehicle natural language systems, in particular to help manage the tension between designing for an engaging dialogue (important for technology acceptance) and designing for an effective dialogue (important to minimise distraction in a driving context).

Keywords: natural language interface; digital assistant; social AIs; driving; simulation; Wizard-of-Oz.

1 Introduction

There is a growing tendency towards humanising interactions with computers and technology, with many common, everyday devices employing natural methods of interaction, such as speech. Advances in speech-recognition technologies, speech synthesis, natural language understanding (NLU) and dialogue management over recent years mean that modern voice-user interfaces (VUIs) are increasingly proficient, apparently no longer stifled by stilted, command-based input, nor plagued by poor speech recognition, incorrect pronunciation or response latency. Nevertheless, human-computer ‘conversation’ – at least that realised by current commercially available technology – is still often restricted to a single exchange, initiated by the human interlocutor, in which a request is made and the computer delivers an isolated response. In contrast, successful bilateral human dialogue is based on the collaborative effort of both interactants to ensure mutual understanding (Buschmeier & Kopp, 2011). Recent efforts to enhance the social element of VUI exchanges has seen modern devices embodied by ‘intelligent’ and ‘socially-aware’ digital assistants (e.g. Siri, Cortana etc.), imbued with human-like ‘personalities’. Nevertheless, if VUIs are to become truly ‘social’ (as appears to be the current desire), there is a need to better understand the role that the host technology plays within the social fabric and context in which it is situated. It is therefore important to understand the language and linguistic characteristics that people expect to employ when addressing and interacting with such a system: this will naturally influence the language and linguistic characteristics that the system should adopt in reply.

1.1 Speech

Speech is a powerful catalyst – it is one of the quintessential markers of humanness and the primary means of social identification amongst humans (Barthes, 1977). From early infancy, humans are able to differentiate speech-like sounds from other sounds in their environment, honing in on the sound of their own mother (Moon, et al., 1993), and use this to maintain a sense of presence, even when their visual field is absent or obscured (Chion, 1999). As humans develop and grow, they rapidly acquire the ability to extract salient, socially-relevant, paralinguistic cues from speech (Barthes, 1977) – even during limited exposure – based on vocal characteristics such as pitch, cadence, speech rate and volume, and use these to provide systematic and highly compelling guidance for determining gender, personality and emotion-specific actions (Nass & Brave, 2005). Extensive research has demonstrated that humans appear to be unable to overcome these evolutionary instincts, and behave towards vocal utterances from a computer in a similar manner, ascribing humanlike characteristics and attending to talking machines as if they were interacting with another human (Nass & Brave, 2005). Using voices and speech as a medium for human-computer interaction has therefore allowed interface designers to exploit this behaviour, with evidence suggesting that different digital ‘personalities’ – created by varying the vocal characteristics and language content of spoken language interfaces – can influence trust, performance, learning and even consumers’ buying habits (Nass & Brave, 2005).

Steady improvements in software integration, personalisation and context awareness over recent decades mean that, in addition to improvements in linguistic proficiency, modern speech-based interfaces are also becoming increasingly ‘intelligent’ and socially aware. This has enabled more pervasive spoken language interfaces, and heralded a shift towards the ‘intelligent personal assistant’ paradigm (Bellegarda, 2014). Indeed, there is a growing number of Artificial Intelligence (AI) assistants – such as Alexa, Cortana and Siri – embodying so-called ‘smart’ technology, that are imbued with human-like personalities (expressed through voice and language): such agents are also beginning to

1 attend to communicative feedback given by their interlocutors and adapt their ongoing and
2 subsequent communicative behaviour to their needs.

3 From a pragmatic point of view, speech-based interfaces also afford ‘hand-free’ and ‘eyes-free’
4 interaction and are therefore often favoured by designers of computers or interfaces intended for
5 deployment in situations already posing high levels of visual and/or manual demand, such as driving.
6 Given the demands already placed on drivers, there is an inimitable attraction and undeniable logic in
7 utilising speech when interacting with in-vehicle technology. Indeed, in-vehicle speech-based
8 interfaces have been commonly explored for mobile phone dialling, navigation destination entry,
9 email processing and music selection tasks while driving; tuning the radio and adjusting climate
10 control have also been examined (Barón & Green, 2006). A common consensus is that using speech-
11 based interfaces in cars improves driving performance (indicated by reduced lane and speed
12 variation), enhances secondary task performance (e.g. fewer navigational errors), reduces workload
13 and improves eye glance behaviour (i.e. less eyes-off-road time), compared to manual interfaces
14 (Barón & Green, 2006). Nevertheless, it is recognised that in-vehicle speech recognition technology
15 and controls do not fully negate the hazards of talking while driving – using an in-vehicle speech
16 interface to dictate text messages or emails, for example, is still likely to contribute to elevated levels
17 of cognitive demand, which has the potential to divert a driver’s cognitive attention away from the
18 driving task (Strayer, et al., 2014).
19
20
21
22
23
24
25

26 **1.2 Talking Cars**

27 Needless to say, it is possible to envisage a future scenario where cars are embodied by autonomous,
28 voice-enabled digital driving assistants, much like current personal mobile devices, that act as a key
29 interface between driver and car, thereby supporting drivers during primary and secondary task
30 execution. Such systems are likely to enable ‘hand-free’ and ‘eyes-free’ operation of vehicle controls,
31 as well as providing all manner of information, such as navigation advice, travel updates, vehicle
32 status warnings, infotainment services etc. It is anticipated that digital driving assistants would also
33 socially engage drivers and keep them apprised of personal information (e.g. news, diary
34 appointments, emails, social media updates etc.). Moreover, one would expect interactions to be
35 initiated by either the human interlocutor or the digital assistant and facilitated utilising free-flowing,
36 conversational language, akin to conversing with a loquacious passenger.
37
38
39
40
41

42 Although the concept of a ‘talking car’ is not new, previous manifestations have been significantly
43 constrained by technological limitations. For example, electronic voice alert systems were installed in
44 Chrysler, Dodge and Nissan cars as early as the 1980s, but only offered a very limited range of verbal
45 alerts and warnings, lacking any discernible ‘intelligence’ or interactivity. More recent interactive
46 examples (e.g. voice activated navigation systems) often required drivers to learn specific command-
47 based vocabulary, but nevertheless have suffered from poor speech recognition and incorrect
48 pronunciation; such devices were also socially and contextually ignorant. However, with more recent
49 technological advancements, the utopian view of an interactive, ‘self-aware’ vehicle entity, which acts
50 as an assistant, advisor, or companion to drivers, is rapidly becoming an engineering reality.
51
52
53
54

55 **1.3 Overview of Study**

56 There is increasing interest in natural language interfaces in vehicles. However, it is unclear how
57 drivers will respond to and interact with an omnipresent spoken vehicle entity that engages and
58 informs drivers and keeps them apprised of the current driving situation as well as their social
59
60
61
62
63
64
65

1 calendar. The demanding nature of the concurrent driving task, in particular, means that interactions
2 may be influenced by factors such as divided attention, high workload and restricted physical posture.

3 Initial in-vehicle VUI enquiries have tended to consider the specificity of in-vehicle speech interfaces,
4 for example, exploring the effect of car voice emotion on driver behaviour, workload or performance
5 (Nass, et al., 2005), or considered certain use-cases or functionality, such as determining the least
6 distracting, or most preferred, voice for delivering navigational advice to drivers (Large & Burnett,
7 2013; Large & Burnett, 2014). Moreover, much of the current research has relied upon self-reported,
8 subjective rating scales to elicit results. In contrast, we utilise a more ecologically valid approach to
9 explore the general character of the ways that drivers speak with the technology while they
10 undertake the routine activities of driving. The aim is to use the findings to determine the role that
11 the system plays within the context in which it is situated and use this to inform future design.

12 To accomplish this, we created an autonomous, ‘highly-intelligent’, digital driving assistant employing
13 a natural language, conversational speech interface, called ‘David’ (which was abbreviated to ‘Vid’ for
14 the study). Participants were invited to drive along a motorway scenario in a medium fidelity driving
15 simulator while using the system (which was delivered using a Wizard-of-Oz approach), and interact
16 with it as they might should such a system exist in the real world. All verbal interactions were
17 subsequently transcribed and analysed using recognised linguistic techniques, such as discourse and
18 conversation analysis, normally reserved for interpersonal investigation.

19 During the analysis, six themes emerged. These are: *formulating responses; turn-taking; back-*
20 *channelling, fillers and hesitation; vague language; mitigating requests and politeness and praise.*
21 Results and analysis are thus presented within these themes throughout the paper. For each theme,
22 we first describe each construct, and then highlight some of the salient features of the language that
23 emerged between Vid and the participants that defined the behaviour. Excerpts from the
24 participants’ interactions with Vid are included, accompanied with relevant literature drawn from the
25 fields of both human communication and human-computer interaction (HCI).

26 Our intention in presenting this work is not to provide a prescriptive model of ‘talking cars’ or
27 comprehensive guidelines; nor do we presume to resolve the theoretical debate concerning the
28 motivation behind social behaviour directed towards computers and technology. Rather, we aim to
29 expose novel, qualitative insights, based on the empirical observation of natural language interactions
30 with an autonomous digital driving assistant, and provide a broad and rich palette for designers of
31 future natural language interfaces, particularly in situations where these are intended for deployment
32 within vehicles or related to other situations of divided attention.

2 Method

The study took place in a medium-fidelity, fixed-based driving simulator at the University of Nottingham, and utilised a standard UK motorway scenario (Figure 1). Twenty-five self-selecting volunteers took part, comprising 19 male and 6 female drivers. All participants were native or fluent English speakers. The age of participants varied – eleven participants were aged between 30 and 40, five between 41 and 50, and nine were over 50; median age range was 41-50 years old – although age was not considered as an independent variable. All participants held a valid driving licence and were experienced and active drivers (mean annual mileage was 7660). Participants were reimbursed with £15 (GBP) of shopping vouchers as compensation for their time.

2.1 Digital Driving Assistant, ‘Vid’

The digital driving assistant, ‘David’ (shortened to ‘Vid’ during the study), was conceptualised as a high-functioning, socially-enabled artificial intelligence. It was described to participants as “a fully operational, prototype natural language system currently under development by a major car manufacturer”, but for the purpose of the investigation, was created using a Wizard-of-Oz approach.

Wizard-of-Oz simulation is a popular, well-established technique, commonly employed to explore future technology concepts. This technique has been used successfully to evaluate proposed design concepts as well as conduct user acceptance studies on finalised interface designs (Dahlbäck, et al., 1993). In a typical Wizard-of-Oz study investigating speech-based technology, a human ‘wizard’ simulates the speech system by delivering verbal prompts and responding to user commands and utterances remotely via a computer or intercom, in a manner such that the user believes that responses are system generated (Kelley, 1984).

During the study, the ‘Vid system’ comprised two Samsung tablet computers, one of which was located within the driving simulator and connected to loudspeakers positioned inside the vehicle. Google Hangouts software was used to provide two-way voice and one-way video chat functionality between the participant and a professional male actor (‘Pablo’), situated out of sight in a nearby laboratory, who had the second Samsung tablet (see Figure 1) – Pablo could see and hear participants, whereas participants could only hear Pablo. A second person accompanied Pablo and was responsible for retrieving facts and figures in real time, as dictated by the course of the conversation (e.g. locating requested music tracks, providing further details about news items, researching answers to specific questions etc.), and supplying these details to Pablo.

[Please insert Figure 1 about here]

Figure 1 – Experimental set up, showing driving simulator with motorway scenario (left) and Vid actor and assistant (right)

During each drive, the Google Hangouts interface was obscured, but the device itself remained in plain sight to drivers in order to provide a tangible source for ‘Vid’ utterances: this misdirection was enhanced by overtly ‘installing’ the system in front of participants. System responses were generated in real time by Pablo, but were guided by a script that incorporated appropriate language and

1 phrasing (avoiding full conversational vernacular), and were delivered in a controlled fashion using a
2 subtle computer inflexion, honed through extensive training and practice sessions, to exceed current
3 state-of-the-art natural language speech interfaces. Pablo was instructed to respond to all driver
4 requests and avoid any clinical, out-of-domain responses, such as “Sorry. I don’t understand”, other
5 than in the event of technical problems.
6

7 Participants undertook 2 drives, each lasting approximately 5 minutes, to explore any differences in
8 their behaviour after returning to a ‘familiar’ system; drivers were engaged in other driving-related
9 activities during the break between Vid drives. Both Vid drives took place on a UK-style motorway,
10 although there were subtle differences between the two scenarios, such as the volume of other road
11 users and general road course (Figure 1). Once participants were established and comfortable within
12 the simulator – and prior to embarking on the first drive – Vid introduced itself using the following
13 script:
14
15

16
17 *“Hello. Let me introduce myself to you. I’m Vid and I’m designed to be your assistant, advisor,*
18 *or companion while you drive. My job is to help you get to where you’re going and provide you*
19 *with useful information and maybe some entertainment along your journey. You can respond*
20 *to me in any way that you like. I can hear you, but I can also see your hand gestures or head*
21 *movements. As your intelligent assistant I can help you with in car tasks and other tasks as*
22 *well. For example, I have access to manage your calendar, reminders, or your favourite*
23 *programs. Please say ‘yes’ to let me know that you can hear and understand me.”*
24
25

26
27 Assuming an affirmative response was received from the participant, Vid said “Great! Thank you” and
28 the participant was instructed to start driving.
29

30 During each drive, Vid initiated conversation using the same opening gambits for each participant (see
31 Table 1), which were presented at approximately the same, predefined times throughout each drive.
32 However, the intercourse naturally developed differently based on participant’s responses. All
33 interjections required a response from participants, who were invited to interact with the system
34 using natural, conversational language, as they might should such a system exist in the real-world.
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

3 Results and Analysis

Each participant's interactions with Vid were transcribed to enable subsequent content and thematic coding, rather than employing strict conversation analysis (CA) conventions (Jefferson, 2004). A content analysis approach was then used to create a comprehensive understanding of the data (Wilkinson, 2011). Although participant demographics varied (age, gender etc.), this was not included as a factor during the analysis as it was not considered as a primary aim of the research.

The subsequent sections describe the six themes that emerged, which encompass the salient features of the language that emerged between Vid and the participants. Extracts from the participants' interactions with Vid are included and accompanied with relevant literature from the fields of both human communication and HCI. It is worth highlighting that the examples do not provide an exhaustive list of all the occurrences in which they have emerged, but rather provide a representation of what exists within the transcribed data.

Table 1 – Examples of Vid's opening statements

<p><i>May I call you by your first name? What is your first name?</i></p> <p><i>I looked at your calendar and you have a meeting upcoming today at 3 o'clock. Would you like me to set a reminder for your meeting?</i></p> <p><i>I've also looked at your to-do list and messages and it looks like you've got a few things to do on your way home this evening. You need to buy milk. Would you like me to set a reminder for you to buy milk?</i></p> <p><i>Would you like to talk about your interests?</i></p> <p><i>Would you like hear the news headlines or sporting headlines?</i></p> <p><i>There is congestion ahead. This may delay you by 5 minutes would you like me to direct you around the congestion? ... OK. I am calculating a reroute</i></p> <p><i>It looks like you'll be late for your meeting. Would you like me to inform them of your new arrival time? Would you like me to leave a message? ... That's been recorded and forwarded</i></p> <p><i>I've noticed you're running low on fuel. Would you like me to locate a fuel station? ... Do you need any particular brand or type?</i></p> <p><i>You are approaching your destination. Would you like me to find an open car park for you? ... Locating an open car park near destination</i></p>

3.1 Formulating Responses

Vid's opening statements (see Table 1) were generally formulated as questions and therefore designed to invite a response from participants. Our first point of enquiry is therefore concerning

1 participants' initial responses. It is evident that there were a wide variety of responses from
 2 participants, which can be categorised as confirmations, semi-committed and polite phrases (see
 3 Table 2 for examples). *Confirmations* typically comprise short phrases or one-word responses – such
 4 as *yes, yeah, great* – which are used to confirm something that has been said, acknowledge that the
 5 comment or request has been understood, or indicate approval and satisfaction within the
 6 interaction.
 7

8 *Semi-committed phrases* are also evident amongst participants' responses, such as *can do* and *why*
 9 *not*. These are less definitive than confirmations, but share a similar function. In contrast to
 10 confirmations, however, semi-committed phrases suggest that alternatives can be undertaken. For
 11 example, in response to “would you like to listen to some music?”, “yes” provides a definitive
 12 acknowledgement from the participant, whereas “can do” suggests that not listening to music, or
 13 listening to something else instead, such as the news headlines, would also be an acceptable
 14 outcome. These semi-committed phrases also share similarities with polite responses. Polite
 15 responses are typically signified by the use of additional features such as praise (“thank you”) (see
 16 Section 3.6), and softening phrases (“please”).
 17

18 Even though the syntactic formation of these utterances were different (e.g. see ‘confirmations’ in
 19 Table 2), the pragmatic meaning behind each of them remains the same. This may pose challenges in
 20 VUIs that are required to understand a variety of responses that can mean the same thing, as well as
 21 understanding and evoking appropriate turn-taking with the participant.
 22
 23
 24
 25
 26
 27
 28

29 **Table 2** – Examples of different response categories provided by participants
 30

31 Confirmations	32 Semi-Committed Phrases	33 Polite Responses
34 <i>Yes</i> 35 <i>Yeah</i> 36 <i>Okey dokey</i> 37 <i>Great</i> 38 <i>Perfect</i> 39 <i>That's right</i>	40 <i>Can do</i> 41 <i>Why not</i>	42 <i>I'll go for the news please</i> 43 <i>That would be very useful,</i> 44 <i>thank you</i>

45 **3.2 Turn-Taking**

46 In order to progress beyond single statements-and-responses, VUIs must be able to deal with multiple
 47 interactions. In human communication, interactions are organised and mediated through the process
 48 of turn-taking. During turn-taking, one party typically speaks at a time and interlocutors cooperate by
 49 alternating the role of speaker and listener, as and when it is deemed appropriate, with the speaker
 50 ‘controlling’ the conversation (Sacks, et al., 1974). Whereas in practice, utterances may actually
 51 overlap to some extent during human conversation, it is evident that the interactions between the
 52 drivers and Vid have clearly marked turns, with participants actively controlling turn-taking during the
 53 conversation with a mix of active and passive approaches. This varied between participants, but also
 54 within the two interactions (i.e. the two drives) that participants had with Vid, with evidence
 55 suggesting that the longer participants interacted with Vid (e.g. when they re-engaged with it during
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65

1 the second drive), the more active and confident they became. For example, during the first
2 interaction between Vid and participant 24 (P24), Vid was the initiator of new turns at talk and the
3 primary nominator of new topics. In the second, and subsequent interactions, P24 takes the initiative
4 after learning more about Vid and its capabilities (extract 1).
5
6

7 Vid: *It looks like you'll be a little late for your meeting. Would you like me to* (1)
8
9 *Inform them of your new arrival time?*

10
11 P24: *How late will I be?*

12
13 Vid: *ETA approximately 7 minutes late*

14
15 P24: *Yeah can you call them please?*

16
17 Vid: *Okay would you like to leave a message?*

18
19 P24: *Yes please. Tell them I'll be 10 minutes late.*
20
21
22

23 After being advised that they will be late for their meeting, P24 requests further information – “How
24 late will I be?” – rather than simply responding to Vid’s statement with a confirmation or denial,
25 suggesting that P24 is comfortable taking the initiative within the conversation. It is also further
26 evidence that they believe that Vid is technically able to process such requests. As the interaction
27 continues, P24 again takes the initiative by asking Vid to contact the other attendees (“Can you call
28 them please?”), and in the final line of this extract, P24 provides both a confirmation of Vid’s prior
29 question, as well as a further request of their own.
30
31

32 This pattern of dialogue suggests that the participant has developed an understanding of Vid’s
33 capabilities and has subsequently modified their language and responses. This is likely to be based on
34 their previous interactions with Vid and is thus comparable to human interactions in which *common*
35 *ground* – i.e. mutual understanding and assumptions between interaction partners – is developed
36 during the course of an interaction (e.g. (Clark, 1996)).
37
38
39

40 The two extracts below show alternative responses to Vid’s declaration that the driver will be late for
41 a meeting.
42
43
44

45 Vid: *It looks like you'll be a little late for your meeting.* (2)
46

47 P18: *Oh dear.*
48
49
50

51 Vid: *It looks like you'll be a little late for your meeting.* (3)
52

53 P22: *Damn.*
54
55
56

57 Although these short responses still provide confirmation that each participant has heard and
58 understood Vid’s message, they also indicate that each participant would like Vid to continue to
59 dictate the course of the conversation. They may also be indicative of a desire for Vid to ‘change the
60
61
62
63
64
65

1 subject', or "topic-shift" (Brown & Yule, 1983), i.e. move from one topic (being late for a meeting) to
2 another (e.g. informing participants about current affairs). It is also possible that, given the content of
3 the statement, i.e. participants were not briefed regarding their agenda, they were displaying apathy,
4 sarcasm, or humour towards the 'news'.
5

6 An alternative explanation is that participants were unsure how to respond given the novelty of the
7 situation (i.e. simulated driving and a prototype natural language system), particularly when they
8 were encountering Vid during the first drive. An additional example is shown in excerpt 4 below. Here,
9 P20 appears to be unsure how to respond when Vid asks, "Would you like to talk about your
10 interests?". P20 avoids providing a direct answer to this vague and broad question and instead
11 responds with their own question in an attempt to clarify the statement, thereby providing Vid with
12 the opportunity to be more specific.
13
14
15
16
17

18 Vid: Would you like to talk about your interests? (4)

19 P20: *participant laughs* Ermm what are you thinking of Vid?
20
21
22
23

24 3.3 Back-Channelling, Fillers and Hesitation

25 Back-channelling reflects the fact that there are two channels of communication operating
26 simultaneously during a conversation – the predominant channel through which the speaker directs
27 primary speech flow, and the secondary channel (or backchannel), through which the listener
28 provides *continuers* or *assessments*, thereby defining their comprehension and/or interest. Such
29 responses are thus received by the current speaker without having to relinquish their turn (Yngve,
30 1970). Back-channelling is used as a way of showing a speaker that you are following what they are
31 saying and understand, often through interjections such as, *I see, yes, OK* and *uh-huh*, although more
32 recently the term has been expanded to include sentence completions, requests for clarification, brief
33 statements, and non-verbal responses. From a VUI system design perspective, back-channelling
34 reflects the fact that both the person who has the turn and his partner may be simultaneously
35 engaged in both speaking and listening, and therefore raises technical questions about how to initiate
36 discussions or topic-shift.
37
38
39
40
41
42

43 There was clear evidence of back-channelling during interactions with Vid. For example, in excerpts 2
44 and 3 (where Vid is the primary speaker), participants use short responses – *damn* and *oh dear* – to
45 confirm that they have heard what has been said and understand the significance of the statement,
46 without attempting to claim the role of speaker themselves (White, 1989). Understanding when
47 people are using backchannels rather than attempting to take the floor at speaking is likely to be a
48 linguistic challenge during these types of interactions.
49
50

51 There was also evidence of hesitation from speakers. During conversation, a pause or hesitation in
52 speech is often masked by *filler words*, which avoid giving the impression of having finished speaking.
53 Fillers are apparently meaningless words, phrases, or sounds, such as *um, uh, er, ah, like, okay, right,*
54 and *you know*, and typically have minimal lexical content, but can play a strategic, syntactic role in an
55 unfolding utterance (Amiridze, et al., 2010). In excerpt 4, P20 uses "ermm" to allow more time to
56 consider the question presented to them and decide upon an appropriate response. A similar
57 example, in which 'hmmm' is used, can be seen in excerpt 5 below:
58
59
60
61
62
63
64
65

1
2 Vid: Would you like to hear the news headlines or sporting headlines? (5)

3
4 P22: Hmmm. Weather please?
5
6

7 In using 'hmmm', P22 displays conscious consideration of the options provided by Vid. Nevertheless,
8 it is interesting that in this case, this ultimately results in a request for a third option – 'weather' – not
9 actually offered by Vid during the initial exchange. Again, this indicates that P22 believes that Vid is
10 capable of processing a different request to those offered.
11
12

13 14 3.4 Vague Language

15 Human speech is rarely always precise and can contain a varied lexicon of vague language (VL). VL is
16 deliberately imprecise language that is used to achieve a wide range of both functional and
17 interpersonal goals, often simultaneously. The description of VL has origins in Channell's seminal work
18 (Channell, 1994). Channell (1994) describes VL as possessing the following characteristics: it can be
19 contrasted with another word or expression which appears to render the same proposition, and is
20 "purposely and unabashedly vague". VL occurs within a wide variety of contexts, including the
21 workplace (Koester, 2007), healthcare (Adolphs, et al., 2007), and in the classroom (Rowland, 2007),
22 and it is particularly common in speech (Channell, 1994; Cheng & O'Keefe, 2014).
23
24
25
26

27 While there has been significant research attention directed to VL in human communication over the
28 past two decades, there has been relatively little literature regarding VL in HCI, yet VL presents a
29 particular challenge to designers of natural language systems. A notable finding is that there are
30 limiting factors such as the quality of a computer's 'voice' than can affect how VL is perceived in HCI
31 (Clark, et al., 2014). To support authentic natural language exchange in the current study, VL was used
32 by Vid in some situations, with participants often employing similar language in response.
33
34
35
36

37
38 Vid: Okay. Open car park located 0.2 kilometres from destination. Is this acceptable? (6)

39
40 P16: Do you have a closer one?
41
42

43 In excerpt 6, P16 uses 'one' as a vague noun. It may also be considered a deictic reference in that it
44 requires shared contextual information to understand the specific object that 'one' describes (Brown
45 & Yule, 1983). Vague nouns substitute the full description of a noun with a concise alternative. Typical
46 examples include words such as *thing*, *thingamy* or *whatsit*. They may also be referred to as "general"
47 or "dummy nouns" (Halliday and Hasan, 2014), as well as "placeholders" (Channell, 1994). This
48 provides speakers with the ability to be efficient with their language use, and avoid repetition of the
49 full noun referent during an interaction.
50
51
52

53 In the above example, P16 substitutes the full noun *car-park* with *one*. This extract indicates that P16
54 believes that Vid will interpret this particular vague reference successfully through an understanding
55 of the context and the previous referents, even though there is another, more recently used noun:
56 *destination*. Another example (excerpt 7) shows Vid using the same vague noun, *one*, which is then
57 repeated by the participant. In this case, the entire phrase *find one nearer* is repeated by P20, rather
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

than the vague noun *one*, demonstrating that participants are also able to hear and understand instances of VL used by Vid, as well as employ VL themselves.

Vid: *BP fuel station located 51.3km. I can find one nearer, but not BP.* (7)

P20: *Yeah, please find one nearer.*

3.5 Mitigating Requests

One of the social functions of VL is to soften expressions and reduce the impression of a speaker being too authoritative or direct (McCarthy & Carter, 2006). There are many overlaps between VL and hedging – lexical items that alter the truth conditions of a statement by attributing “fuzziness” to it and making it less precise (Lakoff, 1973). Such approaches to conversation can help to mitigate the utterances a speaker makes towards a listener. A speaker can also avoid imposing themselves on a listener by using interactional strategies, such as hedging. These linguistic tools can have further overlap into the concept of negative politeness strategies (Brown & Levinson, 1987). These mitigate the imposition a speaker has on the listener’s socially presented self-image or face (Goffman, 1967; Goffman, 2002).

Vid: *Would you like to hear some music by Jimi Hendrix?* (8)

P10: *Erm*

Vid: *Or The Doors?*

P10: *Let’s have a little bit of The Doors.*

[Vid then proceeds to play the music specified by P10].

In this example, P10 uses the hedged phrase *a little bit*, in response to Vid’s question regarding music choice. This can also be seen as a further subcategory of hedging referred to as an “adaptor” (Prince, et al., 1982). This lexical choice attaches vagueness to the request and reduces the class membership of listening to (or ‘having a little bit of’) The Doors. It also signifies to Vid that while P10 is satisfied with listening to The Doors for now, they may change their opinion of the music choice soon. P10 also prefaces their answer with *let’s have* which signals an inclusion of Vid into the choice, and not just P10 themselves.

Vid: *Okay. What would you like me to say?* (9)

P15: *Just say I will be 5 minutes late. Apologies.*

Prior to excerpt 9, Vid has notified P15 that they will be late for a meeting and asked whether they would like to leave a message to inform the other attendees (see Table 1). P15 uses the hedge word *just* in a different manner to the request mitigation seen previously. In this case, the word *just*

1 functions somewhat like a *minimiser* – a lexical item that can reduce the assertiveness of an utterance
2 as well as reduce its perceived difficulty (Clark, et al., 2016). However, P15 mitigates the complexity of
3 their intended response to be delivered to other meeting attendees, rather than being concerned
4 with the perceived difficulty of the utterance. However, P15 is unaware that Vid will record that
5 message and apparently does not expect it to do so. The hedge is perhaps targeted towards both
6 points; it also indicates that the task directed at Vid is a minimal one.
7

8 Mitigation may also take the form of other politeness strategies such as showing gratitude (thank
9 you), apologising (sorry), attending to needs and wants (would you like me to...) and asking
10 permission (may I...) (Brown & Levinson, 1987).
11
12
13

14 Vid: *May I call you by your first name?* (10)

15 P20: *Yes, you may.*

16
17
18
19
20 Vid: *It looks like you'll be late for your meeting. Would you like me to* (11)
21 *inform them of your new arrival time?*
22

23 P20: *Sorry can you say that again Vid I couldn't hear you.*
24
25
26

27
28 P22: *Yes please and let me know how close it is if you can* (12)
29
30
31

32 In excerpt 10, Vid uses a polite request for permission (Leech, 2014) in asking P20 permission whether
33 they can be called by their first name. P20 responds in kind with the use of repetition. “Yes, you may”
34 also an example of using repetition – an integral part of communication, creativity and interpersonal
35 relationships (Carter, 2004). Whether repetition could be exploited in some manner for fostering
36 interpersonal relationships in these settings is unclear, but it is clearly a linguistic tool that participants
37 were willing to employ. In excerpts 11 and 12, both participants mitigate their own requests towards
38 Vid. P20 offers an apology at the beginning of their utterance before continuing with the request,
39 followed by an explanation of why they had to repeat themselves. P22 offers both *please*, at the
40 beginning of the utterance, as well as an *if you can* at the end. This indicates that P22 may anticipate
41 that Vid may not be able to carry out their request, as well as understanding that there may be
42 limitations in the interface’s capabilities.
43
44
45
46
47

48 **3.6 Politeness and Praise**

49

50 Mitigating requests (discussed in Section 3.4) are one aspect in the spectrum of politeness. Previous
51 examples have shown that participants are willing to employ negative politeness strategies in their
52 communication with Vid, such as when they mitigate their requests. There are also examples of
53 participants using positive politeness strategies in their interactions. These strategies can be used to
54 address an interlocutor’s desires and wants. Examples of these strategies include complementing,
55 joking, and showing solidarity (Brown & Levinson, 1987). There were examples of these too in the
56 interaction data with Vid, as seen below.
57
58
59
60
61
62
63
64
65

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Vid: Radio 6. Signal strength good. Connection to Lauren Laverne. (13)

P22: Good work [participant is smiling].

P22 offers praise to Vid for undertaking and successfully completing the task of changing the radio station. The participant's turn at talking may also signify that the issue of selecting a radio station and establishing a good connection has been resolved and they are happy to end this particular discussion topic, while not intending to provide praise *per se*. While the choice of language may not be an attempt as such to praise Vid, this example indicates that the linguistic choices made by participants in these types of interactions overlap with social communication norms associated with human conversation. For designers of VUI systems, this may represent a challenge in understanding the pragmatic function of these utterances, and how best to formulate appropriate responses to them.

4 Discussion

Traditionally, the majority of the research and development attention afforded to human-machine voice user interfaces (VUIs) has been focussed on the computer's ability to disassemble, process and interpret isolated statements made by the human interlocutor. However, as interfaces become more socially situated, developing high-functioning, conversational VUIs demands more than just the technical capability to respond to routine language statements. It is also important to understand the role that the system plays and the social context in which it is situated – this will influence the language and linguistic characteristics that people employ when addressing and interacting with the system, which in turn shapes the language and linguistic characteristics that the system should adopt in reply.

To explore this, we observed how people interacted with a high-functioning, conversational interface (realised using Wizard-of-Oz prototyping), and analysed the language and linguistic behaviours that people adopted. By applying content analysis to the transcribed interactions, the following six themes were revealed, which define the key aspects of participants' linguistic behaviour:

1. Responses were formulated as confirmations, semi-committed and polite phrases, suggesting that VUIs are required to understand a variety of responses that can mean the same thing.
2. During the interactions, participants actively controlled *turn-taking* during the conversation, with a mix of active and passive approaches. Such behaviour needs to be recognised and respected by a VUI system.
3. There was evidence of *back-channelling*, *fillers* and *hesitation* during interactions with Vid. From a system design perspective, such responses must be differentiated from other utterances.
4. *Vague language* was employed by both participants and Vid during interactions. VL presents a particular challenge to designers of natural language systems given its intentional 'fuzziness'.
5. Requests were often *mitigated* using hedging words and expressions, such as *just* and *a little bit*; these also need to be understood by a VUI system.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

6. There were numerous examples of *politeness* and *praise* directed towards Vid. A VUI system needs to respond appropriately to such comments.

The six themes demonstrate a clear departure from regimented ‘spoken interactions’, traditionally associated with VUIs, which have been limited to strict turns at speaking and listening using routine statements (Buschmeier & Kopp, 2011). Instead, interactions were characterised by the continuous exchange of information – using a variety of techniques and language – through which human and ‘computer’ collaborated in order to mutually coordinate and establish shared beliefs and understanding (i.e. ‘natural dialogue’) (Clark, 1996). Participating actively beyond the mere exchange of turns is a prerequisite for robust and efficient dialogue between human partners. Individually, the observed behaviours provide strong evidence that people expect to interact with a digital driving assistant in a similar manner: considered together, the evidence is highly compelling.

It is important to recognise that other linguistic behaviours and techniques – that do not necessarily feature within our themes – may be observed between human dialogue partners. However, the absence of such behaviour does not diminish the strength of our findings. Instead, it is suggested that it actually draws attention to the particular aspects of human speech (that were identified), which can be considered as most applicable during ‘natural language’ interactions with a digital agent in the context of driving.

From a design perspective, the findings present a number of challenges. Indeed, if users expect to speak with a digital assistant as they would another human being – as appears to be the case in a driving context – the digital assistant must respond in like. This suggests that digital agents should understand and even adopt the idiosyncrasies and complexities of human speech, such as back-channelling, fillers, hedging words etc., identified here.

Furthermore, digital assistants have been traditionally designed with low status – based on the subordinate roles that they play and the work that they do – and this has been reflected by the language use patterns and gender (Hannon, 2016). However, it is evident from the current investigation that participants apparently afforded equal social status to Vid, crediting it with high-level cognitive reasoning and social awareness – in a driving context, at least, language patterns and linguistic behaviour should therefore reflect this new, elevated status. This can be accomplished by the selective emphasis of specific language traits and language usage patterns, e.g. eliminating gratuitous ‘I’-words and enhancing the frequency of high-level cognitive-processing words (such as: *think, wonder, because* etc.) (Hannon, 2016; Pennebaker, 2012).

In a driving context, however, there is an additional requirement to ensure that interactions do not distract drivers, and that messages are efficient and effective in delivering the information that is required to support the primary task of driving. Consequently, there is also a need to manage the tension between designing for an *engaging* dialogue (important for acceptance of the technology) and designing for an *effective* dialogue (important to minimise distraction in a driving context). Understanding the linguistic elements revealed in this study, and how these apply in different situations, may assist in managing this tension. For example, results suggest that drivers would expect a digital driving assistant to act as if they were conversing with another human being. Thus, a VUI that behaves in this manner (e.g. utilising the language patterns and linguistic behaviour revealed herein) is likely to promote high acceptance of the host technology. However, in certain situations (e.g. during the provision of navigational instructions), some of the aspects highlighted (e.g. the use of vague language), may be inappropriate, and thus, should be avoided. Moreover, an informal, chatty driving

1 assistant has the potential to distract drivers in situations of high workload (congested urban
2 environment, inclement weather etc.). Therefore, understanding which linguistic elements to use
3 (and which to avoid) in different situations is important. Further investigations should consider these
4 in greater depth.

5
6 It is also worth reflecting on the Wizard-of-Oz methodology used during the study. This highly novel
7 approach provided an excellent platform to observe and comment upon the behaviour of drivers
8 engaged with a fully-functional, natural-language digital driving assistant. Such investigations are
9 predicated on the well-established school of thought that empirical observation is the best way to
10 explore systems for which no complete *a priori* model exists. A particular advantage (in the current
11 context) is that utterances and interactions that fell outside of the well-defined domains of current
12 talking technology could be satisfactorily dealt with, meaning that clinical out-of-domain responses,
13 such as, “sorry. I don’t understand”, or more novel alternatives (Ameixa, et al., 2014), could be
14 avoided – this requires a highly-skilled ‘wizard’. In addition, given the complexities of human speech,
15 it is feasible that other factors, such as subtle paralinguistic cues hitherto yet to be uncovered, or the
16 extreme proficiency of the ‘artificial intelligence’ (i.e. it could not be fooled), could reveal to
17 participants that the system was actually embodied by another human. To combat these potential
18 concerns, our wizard, Pablo, had received extensive training and experience in delivering high-
19 functioning AI responses prior to embodying the digital assistant, and had also conducted several
20 similar research studies previously. We were also careful to uphold the misdirection throughout the
21 study (e.g. ensuring that the technology used to communicate between ‘Wizard’ and participant was
22 robust and appeared ‘authentic’). Indeed, although anecdotal, it is worth noting that all participants
23 expressed genuine surprise when they were informed that they had been interacting with another
24 human-being rather than the technology as described, following completion of the study.

25
26 Provided the approach is well-managed, Wizard-of-Oz methodology is therefore capable of providing
27 a more ecologically valid statement than other investigations into ‘novel’ technology, which may
28 instead rely upon isolated exposure or self-reported, subjective rating scales to elicit results. For the
29 current research, conclusions can therefore be drawn with confidence regarding the likely ways that
30 drivers will interact with a future natural language system, and these can be used to reliably inform
31 the design and development of such systems. This approach is therefore of relevance to other
32 researchers working in the field of voice interaction and the study is also likely to be of interest to the
33 larger ‘user experience’ (UX) community.

34
35 As a final note, it is worth highlighting that the current investigation, by design, aimed to explore
36 language use during interactions, and the transcription conventions were guided by this aim. A more
37 detailed record, utilising stricter CA techniques, may have revealed other linguistic behaviours, such
38 as participants consciously adjusting the pace or cadence of their speech or delivering their
39 utterances in a more ‘neutral’ tone (e.g. to avoid a strong accent). On first glance, it appears
40 impossible to comment on such behaviour from the data as presented. Nevertheless, transcriptions
41 were made with the support of video recordings and transcribers were specifically asked to record
42 any additional, noteworthy behaviours or responses (e.g. excerpt 4 indicates that P20 ‘laughed’). It is
43 therefore likely that any significant changes in participants’ behaviour (speaking slowly, neutralising
44 their tone) would have been highlighted, had they occurred. Moreover, one would expect any such
45 behavioural adaptations to also influence the language and linguistic constructs used. Therefore, from
46 the nature of the language used by participants, and the lack of any contradictory evidence, we
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

1 remain confident that participants behaved in a ‘normal’, unaffected manner when conversing with
2 Vid; moreover, they spoke as if speaking to another human-being.
3
4

5 **Conclusions and Future Work**

7 The way that humans interact with computers – and the role that they play in everyday life – is
8 fundamentally changing. In particular, the current research suggests that people are prepared to
9 afford equal social status to a digital assistant in a driving context and interact with it (or ‘him’?) as if
10 conversing with another human. If we are therefore to imagine a digital assistant as a human
11 interlocutor – as seems to be the desire – artificial intelligences must become more cognitive and
12 social in the work they do and the way that they interact. The current study has revealed a number of
13 key areas of language usage that can assist in accomplishing this goal in a driving context, but further
14 work is needed. In particular, future investigations should consider how to manage the tension
15 between designing for an *engaging* dialogue – in line with people’s ‘social’ expectations, and
16 therefore important for the acceptance of natural language interfaces in vehicles – and designing for
17 an *effective* dialogue, given the other demands already placed on drivers.
18
19
20
21
22
23
24

25 **6 Acknowledgements**

27 The research was conducted in collaboration with Jaguar Land Rover Research and the authors
28 gratefully acknowledge their support. We would also like to thank Vicki Antrobus, Ayse Eren, George
29 Burnett and our wizard, Pablo Raybould, without whom the research would not have been possible.
30 Particular thanks also go to linguistic lotharios, Leigh Clark and Annie Quandt, whose insight and
31 analysis made this paper possible.
32
33
34
35
36

37 **7 References**

- 39
40 Adolphs, S., Atkins, S. & Harvey, K., 2007. Caught between professional requirements and
41 interpersonal needs: Vague language in healthcare contexts. In: *Vague language explored*. s.l.:s.n., pp.
42 62-78.
43
44 Ameixa, D., Coheur, L., Fialho, P. & Quaresma, P., 2014. *Luke, I am your father: dealing with out-of-*
45 *domain requests by using movies subtitles*. s.l., Springer International Publishing, pp. 13-21.
46
47 Amiridze, N., Davis, B. & Maclagan, M., 2010. *Fillers, pauses and placeholders*. s.l.:John Benjamins
48 Publishing.
49
50 Barthes, R., 1977. *Image, Music, Text*. New York: Hill & Wang.
51
52 Barón, A. & Green, P., 2006. *Safety and usability of speech interfaces for in-vehicle tasks while driving:*
53 *A brief literature review*, s.l.: s.n.
54
55 Bellegarda, J., 2014. Spoken language understanding for natural interaction: The siri experience.
56 *Natural Interaction with Robots, Knowbots and Smartphones*, pp. 3-14.
57
58 Brown, G. & Yule, G., 1983. *Discourse analysis*. s.l.:Cambridge University Press.
59
60
61
62
63
64
65

- 1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
- Brown, P. & Levinson, S. C., 1987. *Politeness: Some universals in language usage*. s.l.:Cambridge University Press.
- Buschmeier, H. & Kopp, S., 2011. *Towards conversational agents that attend to and adapt to communicative user feedback*. s.l., Springer Berlin Heidelberg, pp. 169-182.
- Carter, R., 2004. *Language and Creativity: The Art of Common Talk.* s.l.:Routledge.
- Channell, J., 1994. *Vague Language*. s.l.:s.n.
- Cheng, W. & O’Keefe, A., 2014. *Vagueness. Corpus Pragmatics: A Handbook*. s.l.:s.n.
- Chion, M., 1999. *The Voice in Cinema*. s.l.:Columbia University Press.
- Clark, H. H., 1996. *Using Language*. s.l.:Cambridge University Press.
- Clark, L. M. H. et al., 2014. *Potential of imprecision: exploring vague language in agent instructors*. Tsukuba, Japan, ACM.
- Clark, L., Ofemile, A., Adolphs, S. & Rodden, T., 2016. A Multimodal Approach to Assessing User Experiences with Agent Helpers. *ACM Transactions on Interactive Intelligent Systems (TiiS)*.
- Dahlbäck, N., Jönsson, A. & Ahrenberg., L., 1993. *Wizard-of-Oz Studies - Why and How*. s.l., s.n., pp. 193-200.
- Goffman, E., 1967. *Interaction Ritual: Essays on Face-to-Face Behavior*. s.l.:Anchor Books.
- Goffman, E., 2002. *The presentation of self in everyday life (1959)*. NY: Garden City.
- Hannon, C., 2016. Gender and status in voice user interfaces. *interactions*, 23(3), pp. 34-37.
- Jefferson, G., 2004. Glossary of transcript symbols with an introduction. *Pragmatics and Beyond New Series*, Volume 125, pp. 13-34.
- Kelley, J., 1984. An iterative design methodology for user-friendly natural language office information applications. *Transactions on Information Systems (TOIS)*, 2(1), pp. 26-41.
- Koester, A., 2007. About twelve thousand or so’: Vagueness in North American and UK offices. In: *Vague language explored*. s.l.:s.n., pp. 40-61.
- Lakoff, G., 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of philosophical logic*, Volume 2, pp. 458-508.
- Large, D. & Burnett, G., 2013. Drivers’ preferences and emotional responses to satellite navigation voices. *International Journal of Vehicle Noise and Vibration*, 9(1-2), pp. 28-46.
- Large, D. & Burnett, G., 2014. The effect of different navigation voices on trust and attention while using in-vehicle navigation systems. *Journal of safety research*, Volume 49, p. 69.
- Leech, G., 2014. *The pragmatics of politeness*. USA: Oxford University Press.
- McCarthy, M. & Carter, R., 2006. *as visible patterns of interaction. Explorations in corpus linguistics*. s.l.:s.n.
- Moon, C., Cooper, R. & Fifer, W., 1993. Two-day-olds prefer their native language. *Infant behavior and development*, 16(4), pp. 495-500.

1 Nass, C. & Brave, S., 2005. *Wired for speech: How voice activates and advances the human-computer*
2 *relationship*. Cambridge: MIT press.

3 Nass, C. et al., 2005. *Improving automotive safety by pairing driver emotion and car voice emotion*. s.l.,
4 ACM, pp. 1973-1976.
5

6 Pennebaker, J., 2012. *The Secret Life of Pronouns: What Our Words Say About Us*. s.l.:Bloomsbury
7 Press.
8

9 Prince, E., Frader, J. & Bosk, C., 1982. On hedging in physician- physician discourse. *Linguistics and the*
10 *Professions*, Volume 8, pp. 83-97.
11

12 Rowland, T., 2007. Well maybe not exactly, but it's around fifty basically?': Vague language in
13 mathematics classrooms. In: *Vague language explored*. s.l.:s.n., pp. 79-96.
14
15

16 Sacks, H., Schegloff, E. & Jefferson, G., 1974. A simplest systematics for the organization of turn-taking
17 for conversation. In: *language*. s.l.:s.n., pp. 696-735.
18

19 Strayer, D. et al., 2014. Measuring cognitive distraction in the automobile II: Assessing in-vehicle
20 voice-based interactive technologies. *Accident Analysis & Prevention*, Volume 372, p. 379.
21

22 White, S., 1989. Backchannels across cultures: A study of Americans and Japanese. *Language in*
23 *society*, 18(01), pp. 59-76.
24

25 Wilkinson, S., 2011. Analysing focus group data. In: D. Silverman, ed. *Qualitative research (3rd ed.)*.
26 s.l.:Sage, pp. 168-184.
27
28

29 Yngve, V., 1970. *On getting a word in edgewise*. s.l., s.n., pp. 567-578.
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65

Figure 1
[Click here to download high resolution image](#)

