A Multimodal Approach to Assessing User Experiences with Agent Helpers

LEIGH CLARK, University of Nottingham
ABDULMALIK OFEMILE, University of Nottingham
SVENJA ADOLPHS, University of Nottingham
TOM RODDEN, University of Nottingham

The study of agent helpers using linguistic strategies such as vague language and politeness has often come across obstacles. One of these is the quality of the agent's voice and its lack of appropriate fit for using these strategies. The first approach of this paper compares human vs. synthesised voices in agents using vague language. This approach analyses the 60,000-word text corpus of participant interviews to investigate the differences of user attitudes towards the agents, their voices and their use of vague language. It discovers that while the acceptance of vague language is still met with resistance in agent instructors, using a human voice yields more positive results than the synthesised alternatives. The second approach in this paper discusses the development of a novel multimodal corpus of video and text data to create multiple analyses of human-agent interaction in agent-instructed assembly tasks. The second approach analyses user spontaneous facial actions and gestures during their interaction in the tasks. It found that agents are able to elicit these facial actions and gestures and posits that further analysis of this nonverbal feedback may help to create a more adaptive agent. Finally, the approaches used in this paper suggest these can contribute to furthering the understanding of what it means to interact with software agents.

• **Human-centered computing~Human computer interaction (HCI) • Human-centered computing~User studies**
Additional Key Words and Phrases: human-agent interaction, instruction giving, vague language, gestures, facial actions, emotions

ACM Reference Format:

1. INTRODUCTION

The source of many interactions with agents in today's world involves them instructing or advising humans. Examples include those used in satellite navigation systems, intelligent personal assistants on smartphones, and automated checkout systems in supermarkets. How users respond to being instructed by agents as they escape from these familiar boundaries remains a salient topic of investigation [Jennings et al. 2014]. Advising others in human interaction can be a delicate affair as situations are created in which speakers are required to use a variety of communicative strategies to maintain their relationships with one another. There is little information on how these strategies could be used by agents. Some potential has been shown in using the linguistic phenomena of politeness in human-robot interaction.

This work is supported by the Engineering and Physical Sciences Research Council, under grant EP/I011587/1.

Author's addresses: L. Clark, School of Computer Science, University of Nottingham; A. Ofemile, School of English, University of Nottingham; S. Adolphs, School of English, University of Nottingham; T. Rodden, School of Computer Science, University of Nottingham;.

It has been argued that using politeness can help improve user perception of some of their characteristics for example [Torrey et al. 2013], but a number of obstacles exist, such as their appearance and interaction distance [Strait et al. 2014]. Similarly, agents using vague language were observed to have potential in instructing users in assembly tasks [Clark et al. 2014], but again there have been shortcomings in areas including the quality of the agent's voice and the particular choice of lexis used.

These studies have shown that employing communicative strategies in these types of agents needs to be handled with care. In these studies, there has been some success in using politeness to improve users' perceptions of artificial helpers. Users' perceptions were evaluated with quantitative and statistical measure [Torrey et al. 2013; Strait et al. 2014]. However, they do not always provide the full picture as to *why* users perceive helpers in these ways. One of the aims of this paper is to combine existing methods of analysing users' perceptions with a richer qualitative understanding as to why these perceptions occur.

The agent presented in this paper uses speech as the primary mode of communication, much like the agents mentioned previously. This presents unique challenges in understanding the effects of spoken discourse in human-agent interaction, as speech contains a wealth of interactional complexities that build and maintain the way people see each other in terms of power, identity, and personality [Goffman 1959; 1967; Cameron 2001; Coulthard 1992].

This paper aims to compare how people react when taking instructions from agents using vague language in synthesised and human voices. One aspect of the paper focuses on users' perceptions of the agent while the other focuses on spontaneous nonverbal reaction displayed by the users during interaction. In order to develop an understanding of how users respond to agents using vague language and why, a two-way mixed methods approach is presented based on previous vague language research in this field [Clark et al. 2014]. Forty-eight participants were invited to assemble two different Lego models in two separate tasks. A multimodal corpus was created from these interactions.

The paper is structured as follows. The first section outlines the related work and the motivation behind the study (Section 2) before discussing the experiment design (Section 3). The results comprising user perception of vague agents and user nonverbal feedback in interaction are presented in Sections 4 and Section 5 respectively. Section 6 discusses these results, highlights some of the implications of the findings, as well as suggestions for designers of advice giving systems. Section 7 describes the conclusions of the paper.

2. RELATED WORK

2.1 Communicative Strategies in Agent Instructors

As discussed in the introduction, there are many agents that instruct humans in a variety of contexts. Navigation systems are one increasingly common example [Axon et al. 2012]. These include map-based applications in smartphones and satellite navigation systems in cars, both of which direct people using a combination of verbal and visual instructions. Other examples of agent instructors include automated checkout systems in supermarkets [Orel and Kara 2014], and telephone based spoken dialogue systems that have been a mainstay in society for over a decade now [Nass and Moon 2000]. Agents that provide guidance like this may not just be referred to as instructors; they may also be referred to as advice givers or as "helpers" [Torrey et al. 2013; Strait et al. 2014]. Despite the differences in nomenclature, the names all refer to software systems that present information to their users, so the users can complete a task or achieve a goal. In this paper the term*s agents* and *instructors* are used. The agent presented in this paper (Section 3) may not be intelligent or autonomous as such [Wooldridge and Jennings 1995], but still provides users with information to complete a goal, and creates an interaction with a non-human advice giving system.

One of the aims of this paper is to understand how instructions are perceived in Human-Agent Interaction (HAI), when the language used to instruct users differs from the typically direct linguistic style normally associated with agents [Clark et al. 2014]. The vague language presented in this paper is normally associated with a variety of contexts in human-human interaction (HHI).

It is fairly established in the field of Human Computer Interaction (HCI) that people often treat computers as social actors. The Computers as Social Actors (CASA) paradigm and Media Equation posited that in interaction people treat computers as they would do other people [Nass et al. 1994; Reeves and Nass 1996]. One of the key findings of these theories was that HCI is profoundly social and the rules underpinning the interaction somewhat mimic the social interactions in HHI. Consequently, many of the theories from research in psychology, sociology, and other similar fields are also relevant to study in HCI.

The CASA paradigm was expanded upon to suggest that computers can be perceived by users to have personalities similar to humans [Nass et al. 1995]. Even small changes in computers' personalities can elicit social behaviours from their users [Lee 2010]. Additional research has observed that people can identify themselves as a "teammate" with computers [Nass et al. 1996] and can be flattered by computers, much like they would be with other people [Fogg and Nass 1997]. There have been numerous studies investigating the use of humanlike features into computers, agents, and robots in aspects such as language. This includes politeness [Nass and Moon 2000; Mayer et al. 2006; Wang et al. 2008; Scheutz et al. 2011; Torrey et al. 2013; Clark et al 2014; Strait et al. 2014; de Graaf et al. 2015]. Research also includes users' perceptions towards how an interface's voice of verbal technology should sound [Lee et al. 2000; Nass and Lee 2001; Dahlbäck et al. 2007; Jonsson and Dahlbäck 2011; Tamagawa et al. 2011; Grichkovtsova et al. 2012].

Being the social actors that we are, there is a desire to not infringe upon the personal space and rights of others by asking them to do something, and also to not present ourselves in a negative light [Goffman 1959]. To help mitigate these and attempt to build and maintain a rapport with interlocutors, there are a number of linguistic strategies used to manipulate the potential adverse effects of giving instructions [Brown and Levinson 1987]. These include the use of vague language on the part of the speaker and good listenership on the part of the advisee. Listening, according to Oxford [1993], is the process of receiving aural stimuli and giving meaning to it. And this can occur as part of a social process in interactions with people and computers alike [Pinto et al. 2010]. The listener makes use of context, participants and co-text represented by the listener's knowledge of what will and has been said or the content and type of the discussion like instructions to enrich meaning, make it relevant, and aid the decoding process. The other aspect of co-text relates to the listener's knowledge of the language system, such as verbal and nonverbal components of the language that provide the vehicle for communication.

Despite the similarities in HHI and HCI described by the CASA paradigm and Media Equation, there are no guarantees as to the outcomes of applying human communicative strategies in computer and agent instructors.

Previous research in verbal agent instructors has included its use in a real world game setting [Moran et al. 2013]. The use of politeness strategies [Brown and Levinson 1987] has been incorporated into other agent and artificial helper contexts. This includes using politeness to mitigate instructions in both pedagogical tutoring [Wang et al. 2008; Wang et al. 2010] and task-based scenarios [Torrey et al. 2013, Strait et al. 2014]. One of the key aspects of politeness theory is that it is used in human communication to establish and maintain amiable relationships with one another, also known as saving face [Goffman 1959; 1967]. This includes interactions where speakers are performing requests and instructions. Depending on the type of system being interacted with, the effects of politeness HCI so far have been mixed. The studies cited above have indicated the potential uses for politeness. However, there may be aspects of both the agent and interaction that hinder its success. These include the appearance of the helper and the distance in which it interactions (e.g. whether it is in the same room as a user or not) [Strait et al. 2014].

2.2 Human vs. Synthesised Speech

Another obstacle to consider is the quality of the system's voice. Voice is a strong indicator of identity and can influence the way in which we perceive a speaker [Latinus and Belin 2011]. Previous research on computer generated and human speech has shown that using human speech to create an "advanced voice", for example, can result in greater interaction satisfaction than a "basic" computer voice [Cowan et al. 2012]. This study observed no significant statistical difference between the advanced voice and an actual human partner, however.

Another study, comparing synthesised and human voices, discovered that voice actor recordings were rated as more likeable, conversational, and natural than both amateur human recordings and synthesised voices [Georgila et al. 2012]. However, it was also discovered that a "high-quality general-purpose voice or a good limited-domain voice can perform better than amateur human recordings" [p. 8]. They also argued that although the voice actor remains more preferable, the gap between amateur human recordings and synthesised voices has reached a point where the two may perform equally. Prosodic features – the way in which utterances are delivered such as loudness, pitch and rate of speech – can also affect users' perceptions of artificial voices. In one study it was argued that users should experience more positive affect with computers that had prosodic features closer to their own [Branigan et al. 2010]. Agent voice and specific aspects that make up their voices can affect users' perceptions of agent systems, but there is scarce evidence as to the effects of voice on how an agent's language is perceived.

2.3 Vague Language

The perception of vague language in HAI is also not fully understood. A prior experiment observed the quality of the synthesised voice in an agent instructor was often not seen as a good fit for the vague language, and that improving the quality of the voice may make it more suitable for HAI [Clark et al. 2014].

Vague language is deliberately imprecise talk and one of its uses is to achieve functional and social goals simultaneously [Channell 1994; Cutting 2007]. If a word can be contrasted with another to render the same proposition, it is purposefully vague or arises from intrinsic uncertainty then it is deemed to be vague language [Channell 1994].For example, a student answering a mathematics questions in classroom may respond with, "but it's around 50 basically?" [Rowland 2007]. In this example, the speaker conducts the functional goal of answering a question given by a teacher, while also fulfilling the relational goal of protecting oneself from full commitment to the answer and potential error. The student accomplishes this by being imprecise using "around" and "basically", thus saving face. Vague language is seen a wide array of other contexts such as medical examinations [Adolphs et al. 2007], academic conferences [Trappes-Lomax 2007] and the workplace [Koester 2007].

Vague language is one of many communicative strategies a speaker can use to establish and maintain rapport with their users. Vague language in agent instructions can provides a certain degree of human input. This can be seen in the context of the Lego assembly instructions used in this paper. For example, “twist this socket 90 degrees to the right” and “just twist this socket 90 degrees or so to the right” provide the same basic information but in two different ways. The first option is more definitive and precise whereas the second provides a greater number of potential outputs. These instructions and the functions of vague language used within them are discussed further in 3.1.

Vague language is a common feature of everyday speech [Channell 1994]. Despite the successful use of vague language in HHI, it is not fully understood whether it can also be a useful communication tool for verbal agents to employ, and what obstacles it may face when observed in interaction.

2.4 User Spontaneous Nonverbal Feedback

Approximately 93% of the message to which listeners attend is non-verbal [Huczynski 2004]. Nonverbal behaviour includes facial expressions, hand and arm gestures, postures, positions and various movements of the body, legs, or feet [Mehrabian 1972]. In addition, human nonverbal feedback is expressed spontaneously or deliberately through facial expressions and gestures [Ekman and Friesen 1969a; Kendon 2004].

In human communication, active listeners provide appropriate feedback on their perception and understanding of the meaning of messages using contact, perception, understanding and attitudinal reactions [Allwood 1993]. Attitudinal reactions describe the listener’s willingness and ability to react and respond to speaker utterances and reject or accept the message both verbally and nonverbally. These include emotion, which is defined as a unique process of automatic appraisal influenced by our evolutionary and personal past. This enables us to sense that something important to our welfare is happening and a set of physiological changes and emotional behaviours begin to deal with the situation [Ekman 2003]. Spontaneous facial expressions often agree or align with associative expressions like the voice, gesture or posture and are indicative of attitudes and fluency in communication [Hess and Kleck 1990].

Emotion provides the motivation, arousal, adaptive functions, control and variety necessary for effective social interaction between people and other interlocutors. These includes computer animated agents and robots that bring a social dimension to HCI [Bartlett et al. 2010].

Gesture is described as the use of the hands and other parts of the body for communicative purposes [McNeill 1992]. There are different types but the relevant ones for this paper are representational gestures. These gestures may provide a representation of an aspect of the content of an utterance or some feature of the intended referent [Kendon 2004].

Research indicates that representational gestures can be used to represent the form of task objects and the nature of actions to be used with those objects [Fussell et al. 2004]. In addition, gesture scaffolds conceptual development [Capone and McGregor 2004] during interaction. Following these and for the purpose of this research, assembly gestures are seen as the same as representational gestures because they have the following representation techniques. These include placing as if an object is placed or set down within gesture space and sizing as if hands or fingers indicate a specific size or distance [Lücking et al. 2013].

Spontaneous nonverbal behaviour is generated by biologically given processes that operate automatically and elicit facial and limb muscle reactions, quickly and independent of conscious cognitive processes [Ekman 1992]. They are also unmodulated, more reflexive, and smooth and consist of fewer phases than deliberate or posed behaviour [Hess and Kleck 1990]. For example, contractions of the orbicularis oculi – which raises the cheek and gathers the skin inward from the eye socket – is more frequent in genuine than simulated happiness.

Furthermore, the chemicals released when genuine emotions are expressed are never released when they are faked. Posed or simulated emotions are also less symmetric than spontaneous or genuine emotions [Donato et al. 1999; Ekman et al. 1980]. Their timings of muscular contractions are more irregular [Weiss et al. 1987] and there are indications suggesting that simulated emotions are characterised by missing components [Ekman et al. 1988].

In HCI, human emotions have been used in to measure user frustration when interacting with agents [Miles et al. 2013] and monitor emotion regulation with computers [Klein et al. 2002]. Moreover, computers and agents can identify and analyse faces and gestures rapidly and in real time [Zhu and Ramanan 2012]. In addition, systems already exist that enable agents to acquire speech during interaction with humans to make HCI more natural [Rodemann et al. 2010]. One such example is the Honda humanoid robot ASIMO. Following these, we hold that being able to understand user' emotions may enhance artificial advice givers' abilities to interact naturally in much the same manner as humans do when interacting with one another. Studies in human gestures have been used to develop gesture taxonomies for HCI [Kirk et al. 2005] and to assess the manipulation of objects in HCI [Au 2012]. One of the focal points of this paper (Section 5) is to understand whether spontaneous facial actions and gestures also occur when the human is taking assembly instructions from an agent. This section also determines the functions they both perform, and attempts to find a pathway for producing a corpus that may be useful for future agent designers.

3. EXPERIMENT DESIGN

This paper uses one experiment design with two approaches to data collection and analysis. This section outlines the experiment design, agent development, participant population and procedure. There is no non-vague agent being used in this study, however, as direct language is already observed to be a good fit for synthesised voices [Clark et al. 2014]. The specific methods of data collection and analysis for each approach are also discussed.

The first approach hopes to answer the following research questions:

> RQ1: Is there a difference in how synthesised and human agent voices are rated in regards to specific characteristics of the agent?
> RQ2: Is the vague language accepted more in the human voice agent than in the synthesised voice agents?

The second approach focuses on user nonverbal behaviour during interaction and investigated the following questions:

> RQ3: What gestures do users display when requesting repeats of instructions from the agents and why?
> RQ4: What facial actions do users display when requesting repeats of instructions from agents and why?

3.1 Vague Language Model

The salient features of vague language categorised for use in this study are described in this section. Categories based on previous literature are discussed and presented in context of being used by a verbal agent instructor in Lego assembly tasks.

3.1.1 Hedges. Hedges are lexical items that alter the truth condition of a statement by attributing "fuzziness" to it i.e. utterances are made less definite and precise [Lakoff 1973]. Hedges have different functions depending on the type of being used. Prince *et al.* [Prince et al. 1982] describe two categories of hedges: *shields* and *approximators*. Shields themselves are divided into *plausibility shields* and *attribution shields*. Plausibility shields are phrases that a speaker uses to declare a degree of uncertainty to a statement they are making (e.g. *I think, possibly, as far as I can tell*). Attribution shields deflect responsibility of a statement to someone or something other than the speaker (*it has been said that, according to X*) [Fraser 2010].

Approximators, the other class of hedges, are also subdivided into two categories: *rounders* and *adaptors*. Rounders provide estimations, usually of measurement, and convey a range of values (*approximately fifty metres, about here, around half past ten*). Adaptors create imprecision through the reduction of class membership (*somewhat, sort of, kind of, a little bit*) as opposed to using a definite alternative (see Fig 1).

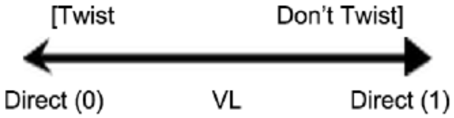


Fig. 1. Vague Language (VL) occupies the space between the two direct alternatives

The directive to twist appears in the context of the Lego assembly instructions used in this paper. One example includes, "give this piece a little bit of a twist". Though twist may arguably be a vague term in itself, it implies some form of rotation of an object from its current position in 3-D space. The adaptor phrase *a little bit* modifies the class membership of the verb twist. The benefit of this is twofold. Firstly, the adaptor helps to mitigate the impact of the instruction, which may otherwise appear assertive and abrupt. Secondly, it opens up a wider set of possibilities for the listener that exists between the direct alternatives, allowing for a larger number of potential outputs. It is for these reasons that they have potential in human-agent interaction and are included in this model.

Rounders that pertain to measurements and value estimations are excluded from the model, as are both classes of shields. An agent instructor using rounders, for example guessing the amount of pieces to pick up, does not suggest competency in its ability to collaborate with a human in an assembly-building context. It is assumed users will have expectations regarding the expertise of the agent and using shields does not promote this either, nor it is the focus of this study.

3.1.2 Discourse Markers. Discourse markers are words or phrases that function primarily as a structuring unit of spoken language [Fraser 1990; Jucker et al. 2003]. Despite containing no grammatical information, discourse markers are common in natural speech [Laserna et al. 2014]. Examples of discourse markers include *now, well, so* and *actually*. Structurally, they can be used as a bridge from one section of information to another, as well as to indicate a change in topic. The ability to structure different turns of information already makes them ideal for an agent delivering assembly instructions, where there will be various stages and sub-stages of building involved. In a humanoid Lego model, for example, the lower body may be seen as one stage, with the feet one of the sub-stages in the assembly process. Discourse markers can help group these together and alert users to a shift in the assembly process.

Discourse markers are not a feature usually discussed in vague language. However, structuring talk is not their only feature. They can also operate as a hedging device by reducing markedness of phrases that may have an effect on a listener, indicate loose or non-literal utterances, and lessen the assertiveness of a speech act [Andersen 1998, Fleischman and Yaguello 2004; Adolphs et al. 2007].

3.1.3 Minimisers. This definition of discourse markers does come with a caveat. Although some can have hedge-like effects, there are some that may be more appropriate than others and blur the lines between discourse markers and hedges. These are described in this paper as minimisers, a term borrowed from describing the use of 'just' as a tension management device in academic presentations [Trappes-Lomax 2007]. They are used in three different forms – *like, basically* and *just*. While discourse markers such as *so* and *now* operate primarily at the beginning of information structures, minimisers appear both at the beginning and mid-sentence. Minimisers can be understood as discourse markers that can function strongly as hedges and vice versa. Minimisers also aim to simultaneously reduce the assertiveness of an instruction whilst reducing the perceived difficulty of the task associated with that instruction. Take the following examples from the Lego instructions:

> *Step 12: Now locate the largest black piece that has seven ball joints. This is the body.*
> *Step 37: So now locate the yellow face piece.*

In the above examples, *now* and *so now* are operating at the start of a new stage in the assembly process. Step 12 introduces the body and Step 37 the face. Compare these to the next examples:

> *Step 13: Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs.*
> *Step 22: Just connect the yellow joints to the socket of each fist.*

In Step 13, *basically* attempts to indicate that the task of finding the narrow end and attaching the ball joints to the leg sockets are not challenging, showing belief in the users' capabilities and minimising the imposition of the instruction. Step 22 is similar and places the subsequent phrases in more positive light.

> *Step 9: So keep these black pieces vertical and just twist each one a little bit 90 degrees or so to the right. These are the legs.*

Step 9 provides a final example. This step begins with the discourse marker *so* but also includes the minimiser *just* in the middle of the sentence. If these were to be interchanged it would look as follows:

> *Step 9: Just keep these black pieces vertical and so twist each one a little bit 90 degrees or so to the right. These are the legs.*

The alternative orientation of Step 9 above lacks the mid-sentence flow provided by *just*, yet there is nothing out of place about it being used at the beginning of a sentence.

3.1.4 Vague Nouns. The final category of the vague language model is vague nouns. Also referred to as "placeholders" [Channell 1994], they substitute the full description of a noun with a concise alternative such as *thing*, *thingamy* or *whatsit*. In terms of potential to achieve relational goals, vague nouns are not expected to contribute towards rapport management or facework as much as other categories. Instead, there is a stronger functional focus on their use in the instructions. Using nouns such as *thing* allow for greater language efficiency, as it reduces the need to repeat noun phrases in full. This prevents the interaction from quickly becoming tedious, which was observed when assembly instructions were designed with too many diagrams [Agrawala et al. 2003]. Similarly, in HCI, having more information than required may be undesirable [Niculescu 2011]. Furthermore, vague nouns can reduce the length of instructions and ensure maximum potential exposure of the agents' speech to the participants.

For the purposes of this research, nouns such as *piece*, *end* and *thing* that are used in the instructions can be said to be vague nouns, in that they operate in a similar manner to *thing* in human interaction. While *piece* and *thing* can be used to one of the constituent parts of the Lego models, *thing* may arguably represent a more open set of potential nouns. It may be the case that most participants equally attribute them both to parts of the model. However, it is believed *piece* is better attributed to the parts of a model assembly. Because of this, it is included in the majority of steps throughout the instructions, whereas *thing* is restricted. *End* is also used in each model's instructions. This noun may not be as vague as the others, as *end* refers to an extremity of one of the constituent model parts (e.g. opposite points on a cylinder).

The model shows there are challenges in clearly defining the categories of vague language and their boundaries. As is often the case, the boundaries are culpable to cross into one another. Importantly however, this gives a more thorough account as to what vague language is, an example of it in use, and its functions. Table 1 gives a concise visual overview of this section and a complete list of items used in this research.

Category	Items in Model	Linguistic Function	Example in Instructions
Adaptors	More or less; a little bit; sort of; a bit; a little; pretty much; or so; somewhat	Hedging instructions: reduce assertiveness; minimise imposition; reduce face threats	So keep these black pieces vertical and just twist each one a little bit of a twist 90 degrees or so to the right
Discourse Markers	So; now	Structure new turns at talk; some hedging	Now pick up the two black pieces with ball joints
Minimisers	Just; like; basically	Structure talk; hedging; reduce perceived task difficulty	Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs
Vague Nouns	Thing; piece; bit	Improve language efficiency;	Just place the big spikes into the holes that are closest to the edge of each piece

Table I. A summary of the vague language model

3.2 Agent Design

The interfaces that provided the forty-seven instructions to participants exist as HTML files. These files are linked to a library of .wav sound files containing the verbal instructions. Three agent voices and two Lego models were used, creating a total of six different interfaces. On the interfaces, participants were provided with the Lego model name they were building (Aquagon or Nex) and the current step of the instructions that they were listening to or had listened to most recently (see Fig 2). Below this information were the two buttons they could interact with. The first was *Next Instruction* (or *Start* for the first step*)*, which moved participants onto the next step of the instructions. The second was *Repeat* (or *Finish* for the last step*)*, which repeated the most recently played step.

The interfaces also functioned as a tool for logging task data, though these capabilities were hidden. The interface logged the time taken for each step, the number of times participants requested for repeats, and the steps repeated. Each interface was given to the participants on a MacBook Pro 10.2.

For the voices of the agent interfaces, two were synthesised and the third was human. The synthesised voices were Cepstral Lawrence (CL) (https://www.cepstral.com) and CereProc Giles (CP) https://www.cereproc.com). The instructions for both models and voices were inputted into a text-to-speech program (Text2SpeechPro) and exported as .wav files. The third voice was a human recording provided by a professional voice actor (VA). The actor was hired from http://voicebunny.com. VA had similar voice characteristics to the two synthesised voices – male, a southern RP English accent, and aged between 45-60 years old. VA was also instructed to sound similar to the synthesised voices in terms of voice quality and cadence of speech.

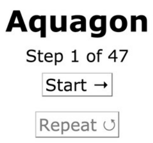


Fig. 2. An example interface that appears on the machine given to participants

3.3 Population

A total of forty-eight native English speakers were recruited for the experiment and reimbursed with a £10 voucher for participating. Twenty-one participants were male (43.8%) and twenty-seven were female (56.2%) and together had a mean age of 24.2 years (SD = 5.56). Three different groups comprised of all the voice pairing combinations were created – CL & CP, CP & VA, and VA & CL. Participants were randomly assigned to one of these groups. There was also balanced in both voice order and model order to consider, which in total created 12 different group iterations. As there were forty-eight participants in total, there was an even counterbalance of four in each of iterations. In total there were 96 interactions with the agents and each voice was involved in 32 of these.

3.4 Procedure & Task

Before each session participants were requested to fill out a short demographics form and were informed that they were to construct two different Lego models under the guidance of spoken instructions from a computer interface. Participants were also informed they had fifteen minutes to get as far as they could in each task. A timer was set so participants could keep track this time limit. After the debriefing the first model was presented and the task began. Once the time limit expired or the model was complete, participants were provided with questionnaires to fill out followed by a brief semi-guided interview. This procedure was repeated once more with the second Lego model.

3.5 Data Collection & Analysis

Each session was recorded from two different angles. A Panasonic HDC-SD900 captured the close up shot of the participant’s face and a Canon Legria HFR306 recorded from the side to capture both the nuances of interaction with the interface and the model assembly (see Fig 3). Although each camera had the capability to record in full high definition, early trials showed the file sizes to be too large for storage. The smaller .mp4 format was used as a substitute without compromising greatly on quality.

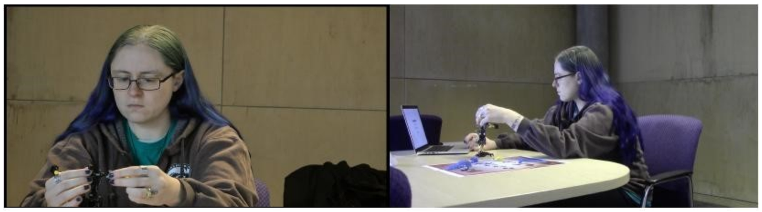


Fig.3. Front and side camera angles showing a participant interacting with the interface and one of the Lego models.

For the first approach a combination of quantitative and qualitative methods was used to assess the data collected from the general experiment design. The questionnaire given to participants following each assembly task contained eighteen questions pertaining to particular characteristics of the agent. These include such as likeability, coherence and human likeness and used five-point Likert-type rankings (1= Strongly Agree; 5= Strongly Disagree). An additional open question on the participants' perceptions of the age of the agent was also included

The qualitative data consisted of the semi-structured interviews from each recording. These were transcribed to create a written corpus of 60,000 words. The video data for each interview was transcribed in CLAN (Computerized Language ANalysis) - software designed specifically to analyse data transcribed in the CHAT format[1]. Each transcription was inputted into NVivo - a qualitative data analysis software package used to collect and analyse text-based and multimedia data. The transcriptions were then coded to identify the attitudes towards vague language and the agent's voice. Quantitative coding focused on coding each transcription according to whether participants had noticed the occurrence of vague language, their overall perception of vague language in the interaction (if observed), and their general attitudes towards the agent's voice. In each interview participants were asked their opinions on the agent's language use, its voice and how the combination of the two functioned together. This helped contribute to their overall attitudes towards these features of the agent. In assessing whether participants observed any vague language, their interview responses were coded depending on if they identified any vague language (either *yes, unsure* or *no*). The coding was as follows:

Yes – Participant describes the vague language in general terms ("filler words", "fluff") or in specific terms using examples from the instructions ("basically", "just").
Unsure – Participant describes the language of the agent but do not refer to either general or specific terms.
No – Participant does not describe the language of the agent in either general or specific terms.

Qualitative coding attempts to address the reasons why users perceived agents in the manner they did. This was achieved by observing and compiling patterns of data from the interview responses.

For the second approach, a multimodal corpus of twenty-four hours of video between participants and the interface recorded, built and analysed. This involves

[1] For information see http://www.childes.talkbank.org/qualitative

the selection of resources that make the corpus representative [Halliday 1991; Jewitt 2009; Lemke 1990]. CLAN and ELAN (EUDICO Linguistic Annotator) were used to build, segment, demarcate, and annotate the corpus for gestures and FACs using adapted schemas, coding systems, and hierarchies [e.g. Feng and O'Halloran 2012; 2013; Ekman and Friesen 1969a; 1969b; Kendon 1980; Wittenburg et al. 2006]. This corpus is kept at the University of Nottingham and accessible with the researchers' permission.

The second approach focuses on the instances in which users repeat instructions, via the agent interface, as a focusing lens for assessing interaction and marked user nonverbal feedback. Carter [2004] holds that repetition is a resource by which interactants create discourse and relationships. In addition, it is a pivotal linguistic meaning-making strategy and resource for individual creativity and interpersonal interaction. By analysing repeats, the second approach aims to assess what can be understood in these instances through the users' FACs and gestures.

In answering the research question, the analysis in the second approach focuses on the basic features of the data gathered because a simple description of what is and what the data shows makes it easy to identify nonverbal feedback [Trochim and Donnelly 2001].

4. IMPACT OF VOICE ON USER PERCEPTION OF VAGUE AGENT INSTRUCTORS

This first approach evaluates whether there are any observable effects on users' perceptions of agents across the three different agent voices (CL, CP and VA). In particular, this approach addresses how users perceive vague language used by these voices, and whether there are any differences in users' responses when comparing the synthesised and voice actor agents.

4.1 Quantitative Results

A one-way between-subjects ANOVA in SPSS was conducted to compare the mean values of each attribute used in the questionnaires across the three agent voices. These were all followed with post-hoc Bonferroni corrections. The ANOVA revealed that there was a significant difference in the likeability of the voice, *F* (2, 93) = 14.77, *p* = < .001. Post-hoc Bonferroni corrections showed that VA was significantly more likeable than CL (*p* = < .001) and CP (*p* = .001). Similar significant differences were observed in how annoying each voice was, *F* (2, 93) = 8.68, *p* = < .001, with VA significantly less annoying than CL (*p* = .001) and CP (*p* = .002). Significant variation was found in how coherent the voices were rated, *F* (2, 93) = 3.43, *p* = .036. VA was rated as significantly more coherent (marked as "Incoherent" in Table 2) than CP (*p* = .033), though no other differences between voices were observed. Ratings of kindness were significant, *F* (2, 93) = 3.36, *p* = .039. Bonferroni corrections, however, revealed no further significant differences between the voices, though the difference between VA and CL was close (*p* = .058). A significant difference was observed in how much each voice enabled participants to complete the task, *F* (2, 93) = 4.24, *p* = .017, with both VA (*p* = .04) and CL (*p* = .04) rated as significantly enabling task completion more than CP. Finally, there was a significant difference observed in how humanlike each voice was rated, *F* (2, 93) = 15.004, *p* = < .001. VA was rated as significantly more humanlike than CL (*p* = < .001) and CP (*p* = < .001), whereas there was no difference observed between CL and CP themselves.

Table 2: Significant differences observed in the characteristics from the questionnaire. Lower mean scores indicate a higher rating for that characteristic.

Voice	Likeable***		Annoying**		Incoherent*		Kind*		Enabled Completion*		Humanlike ***	
	M	SD	M	SD	M	SD	M	SD	M	SD	M	SD
CL	3.31	.965	2.72	1.143	3.47	.950	3.06	.914	2.03	.999	3.88	.976
CP	3.66	.971	2.78	1.008	3.25	.842	3.00	.672	2.69	1.223	3.88	1.238
VA	2.34	1.066	3.69	.965	3.84	.954	2.56	.914	2.03	.861	2.50	1.244
TOTAL	3.10	1.138	3.06	1.122	3.52	.940	2.88	.861	2.25	1.076	3.42	1.319

*p values: * = p < .05; ** = p < .001; *** = p < .00001*
M = Mean; SD = Standard Deviation

4.1.1 Perception of Vague Language and Voice. In the interviews participants were questioned on their observations on the agent's language. Table 3 shows the frequencies of users noticing vague language across the three voices. Note that the total of 94 is the result of missing interview data as previously mentioned.

Table 3. The number of times vague language was noticed or not across each voice condition.

	YES	UNSURE	NO	Total
CL	20	4	9	33
CP	16	3	13	32
VA	15	0	14	29
Total	51	7	36	94

In comparing the frequencies it was found that, in the majority of tasks, participants noticed vague language i.e. they explicitly referred to it during interviews. However, there still remained a significant number of interviews where it was not mentioned. There was a slightly majority in the number of times it was noticed in CL.

Table 4: Frequency of positive, neutral and negative attitudes towards vague language in the voices.

	Positive	Neutral	Negative	Total
CL	1	4	15	20
CP	1	3	11	15
VA	4	8	3	15
Total	6	15	19	50

In assessing attitudes towards vague language   in each voice there is a clear disparity between the numbers in the negative column (Table 4). There were far more instances of vague language being seen as a negative feature of the agent when being used with the synthesised voices. However, despite this favorability there were low numbers in the positive reactions to vague language in all three voices. There were also 15 comments on vague language that were observed to be neutral. The neutral category represents instances where participants explicitly referred to the vague language, but did not display clear positive or negative attitudes towards it (e.g. "I didn't mind it.").

Table 5. Frequency of positive, neutral and negative attitudes towards the three voices in general.

	Positive	Neutral	Negative	Total
CL	1	12	12	25
CP	1	16	12	29
VA	18	8	3	29
Total	20	36	29	83

When analysing the attitudes towards the voices separate from the language, there was a large difference between how the synthesised and human voices were perceived (Table 5). CL and CP had a greater number of both neutral and negative attitudes given in the interviews, whereas VA had a significant majority of the positive attitudes. The somewhat even distribution of neutral and negative attitudes towards CL and CP indicates there is somewhat of a middle ground between neutral and negative where these synthesised voices fall. The overwhelming majority of positive reactions towards VA indicates human voices are still preferable to synthesised voices. However, with VL the differences between the voice types are not as definitive. This suggests there is still some resistance towards accepting VL in this context, even with a voice actor.

4.4 Qualitative Results

Analsying the interview data from the written corpus of interview data produced several patterns or theme. These are presented in this section, with some examples taken from the data, along with a note of which voice condition the quotes originate from.

4.4.1Voice and Language: Disparity and Good Fit. Table 4 indicates that vague language appears to be more suited to VA than the synthesised voices. Those participants that did comment on the vague language specifically, often commented that the two synthesised voices were not congruent with the use of the humanlike nature of vague language:

> P15: "Well it's a bit disconcerting because the voice is like robotic so you wouldn't think it would have like it would be unsure of itself or not quite sure what to say". *(CL)*

> P29: "It sort of made me a little bit uneasy because it was likit was trying to be human but it wasn't human". *(CP)*

The combination of "robotic" voices and "human" vague language contributed to creating unease and a sense disparity for participants. For example, the use of vague language created an agent that could be perceived as unsure in its own advice. Similarly, there was a notable pattern of unease and strangeness arising. Table 4 indicated more positivity towards vague language and VA and the combination of the two being a good fit:

> P28: “I think it can work, because that one worked a lot better. Maybe it just needs to be the right kind of voice for it”. *(VA)*

> P32: “It seemed more like a natural conversation”. *(VA)*

The naturalness of the voice actor and the language combined to alleviated some of the eeriness observed with the synthesised voices. One participant compared the fit of the language in two voices (VA; CL) to roles in the medical professions:

> P19: “It was like the difference between how a doctor *(CL)* talks to you and how the nurse talks to you”. (*VA)*

> P19: “You feel like you want to ask this one *(CL)* more questions because the other one *(VA)* is friendly and clear enough that you can just get on with it, whereas it uses all those words and I’m sat there going I don't understand, say it again a bit slower”.

In this example, P19 perceives VA as a more friendly and approachable nurse figure and CL as a doctor with more expertise. Because of the perceived friendliness of VA they are more able to “get on with it” whereas the more authoritative presence of CL invites itself to more questions, seemingly born out of frustration. The need for CL to be slower in this extract is also an example of the confusion brought about by this synthesised voice’s prosody.

The CP voice also compared with people, though not always favourably:

> P30: “I heard it say like at one point and it was saying totally and like basically and I was like am I talking to a valley girl or something? Am I listening to a valley girl robot? What's going on here? Strange when you're hearing it from something that seems so unhuman”. *(CP)*

In the example above, CP draws a negative comparison to a “valley girl” - a stereotype hailing from California associated with the use of words such as *totally* and *like*. Again this creates a sense of strangeness for the participant as this non-human instructor is speaking in some very human ways.

4.4.2 Social Positioning. The extracts discussed in 4.4.1 may be considered seen as examples of participants placing agents in society. That is to say they are attempting to categorise and position them within their existing frameworks of identity, for both agents and humans, and instances of them overlapping. One of the unique patterns of this that occurred fairly frequently was when participants were followed up on the question regarding the age of the agent. Although many were able to offer up a guess at a figure, this question often brought about a different type of answer:

> P17: “So yeah that's the thing I guess it's almost as if a human has an age as in the years they've lived but a robot has not an expiry date but like a date of production. I wouldn't call that an age”. *(CP)*

> P20: “Erm I said thirty, but to me I didn't actually attribute it to a human voice at all”. *(VA)*

> P43: “Unless it's a human voice with tone you can't judge it”. *(CL)*

It appears that when participants attribute the agent to being more robotic than human, they are less able to give it an age. This notion of having a “date of production” rather than an age can be seen as an example of the out-group social identity that agents are placed in. This is not necessarily a bad thing, but shows a difference in identifying characteristics between the two. Even the voice actor sometimes was attributed to as not being human and an age could not be given to it. In some instances however, either the use of a human voice or the language could change the user’s perception of the age and of their ability to compare it to other people:

> P38: “The terminology as I just said, sort of saying like and just and also the more upbeat, cheery nature, I placed it a lot younger”. *(VA)*

> P40: “I was able to compare it to actual people who might sound like that and their age rather than like a computerized voice” *(VA)*

Much like the good fit discussed earlier, the VA was more likely to be associated with possessing an age. Sometimes this was in because of the language and sometimes it was solely because of the voice. In some cases, participants viewed the agent as younger because of either or both of these features.

4.4.3 Influence of Expectations. In a similar vein to social positioning, there were patterns in participants’ attitudes towards the agent being affected by their expectations of the agent’s speech. Sometimes these were positive or fairly neutral, though these categories were more common with the VA voice:

> P27: “That's the voice that you would expect really in this sort of stuff [instruction giving] so it doesn't bother me”. *(CL)*

> P44: I'd say it was good like; it sounds like the typical voice you kind of hear for that kind of thing, I don't know if you get what I mean. It's just the way he talks”. *(VA)*

Again, much like the other themes presented in this section, negativity was more common in the CL and CP voices:

> P7: “Because you don't expect that [vague language] coming out of machine sort of voice so it sounded a little bit contrived”. *(CL)*

> P37: “Just [a] very standard kind of digital voice kind of, it's like Siri and things like that. It's meant to be calming but it's not really. It feels a bit alien”. *(VA)*

This last example shows the influence that previous agent interactions can have on future ones. It would appear that P37’s prior interactions with Siri, for example, have created a negative perception of both Siri and other similar agents, because of its digital tone of voice. With P07 there was no expectation and perhaps no desire to hear vague language with a more robotic voice. P07 perhaps has some prior experience interacting with machine voices that used a more direct approach in its language, thus creating a negative perception here. Prior interactions with P27 and P44, however, appear to have been different experiences and contributed towards creating a more positive perception in these interactions

5. User Nonverbal Feedback When Agents Repeat Instructions

This section focuses on the subtle ways in which users convey their feelings, as meanings are not determined solely from spoken language. The analysis is ethnographic which aims at developing a 'thick description' [Dörnyei 2007] by providing a detailed, systematic, and rigorous analysis of each marked occurrence. This enables the identification of meanings generated by participants and the emerging trends attached to the behaviour.

In addition, multiple inter-rater agreements ensured reliability and validity [McMillan 2014] of the researcher's interpretation of nonverbal feedback. Five assessors went through six videos and indicated their perceptions of the nonverbal feedback displayed. Inter-rater reliability coefficient correlation for the nonverbal behaviour displayed by the participants indexed by Fleiss kappa is 0.31 while Krippendorff alpha is 0.29. The inter-rater analysis of each assessor with the researcher indicates that assessors 1 and 2 had coefficient correlation of Kappa 0.55 or alpha 0.59 implying a moderate agreement with the researcher. On the other hand, assessors 3, 4 and 5 had a slight agreement with Kappa 0.14 or alpha 0.21. Generally, there is fair agreement between the raters' and the researcher's assessments.

The findings of this approach discuss repeats, emerging gestures, and facial actions specified in RQ3 and RQ4. However, future work may be conducted to test the statistical validity of these basic observations

5.1 Repeats

Repetition is an important aspect of human interaction that is tied to linguistic theories of repairs [Schegloff et al. 1977] politeness [Brown and Levinson 1987], task performance [Kim and Tracy-Ventura 2013], and listenership [Tsuchiya 2013]. Repetition is a pattern that is potentially present in language, and language users have a choice whether or not to form such patterns. Language users also perform the various forms of repetition to project their way of seeing things and create greater mutuality between them [Carter 2004]. A random selection of individual instructions or specific time marking was not used because there was no assurance that either could be adequately covered during interaction. For example, some participants could not cover all the assembly steps within the time limit. A large number of repeated instructions were observed (Table 6), suggesting that repeats constitute a significant aspect of this specific HAI context.

Table 6: Agent instructions repeated

	INSTRUCTIONS				**REPEATS**			
	CP	CL	VA	TOTAL	CP	CL	VA	TOTAL
AQUAGON	47	47	47	141	116	109	75	300
NEX	47	47	47	141	138	117	112	367
TOTAL	**94**	**94**	**94**	**282**	**254**	**226**	**187**	**667**

Table 6 shows that there were 47 instructions for each combination of voice and model. In total, there were 667 repetitions requested by participants. Of these 667 repeats, 55% were in interactions involving the model Nex, while 45% were in interactions with the other model, Aquagon. Additionally, interactions with CP encountered 38% of the repeats, while 34% of these occurred in CL and the remaining 28% in VA.

5.2 Repeat Process

The results also indicate that just as in verbal communication, there is a systematic process for repeating instructions even though the instructor is an agent that cannot hold a two way communication with the user although, the user can repeat instruction as many times as desired (See Fig 4.).

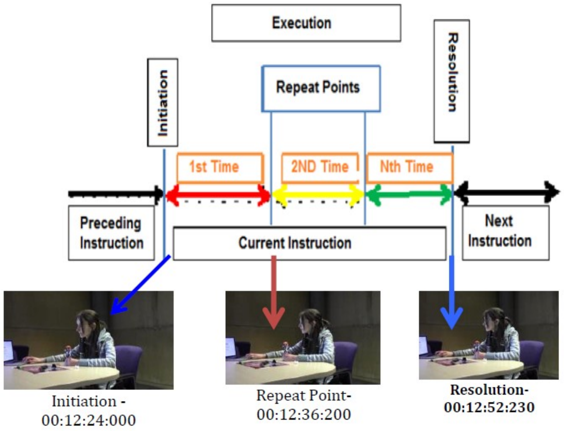


Fig. 4. The repeat process in the assembly tasks.

The study suggests that the process of repeating instructions (Fig 4.) is a discernible, systematic, non-arbitrary and continuous chain of events that has three major stages: initiation, execution and resolution. The repeat process begins with initiation for the first time the instruction is given. The next stage is execution, in which the user executes the instruction given by the agent, says or presses the "Repeat". At the intervals where this button is pressed, these are referred to as "repeat points"**.** The repeat point is achieved when the agent finishes relaying the current instruction, the user requests a repeat, and the agent executes the repeated instructions. Resolution is the point at which the user asks for the "next instruction" after executing the current instruction. In this figure, the repeat process is initiated at 00:12:24:000 by the participant when the preceding instruction (23) has been executed. The Repeat Point (RRP) occurs at 00:12:36:200 when the participant asks the agent to repeat instruction 24. Resolution takes place at 00:12:52:230 when the participant asks for the next instruction which commences at 00:12:53:480.

Repeats are discernible in whatever form they occur for example, the user pressing the "Repeat" button on the interface makes such instruction heard by the ear of the user. Repeats are organised, well ordered, and planned actions by the user to achieve a given aim. For example, initiation comes before execution not vice-versa.

5.3 Typology of Repeats

The following table (Table 7) presents the typology of repeats that have emerged from the interactions with the agent instructors. The table presents the nine types of repeats that emerged from the study. The users perform repeats for different functions, which are explained in the subsequent paragraphs of this section.

Table 7: Typology of repeats

TYPE	CODE	FUNCTION
Clarification	R1	Self-correction
Clarification	R2	Reduce confusion and clear doubts. Usually occurs after listening to an instruction
Confirmation	R3	Affirms or disproves interpretations of agent instructions or assembly processes, and helps self-assurance. Usually occurs after listening to instructions or executing assembly processes
Composite	R4	Multiple repeats for different for different purposes, with a focus on executing one step in the assembly
Demarcation	R5	Used to break up one step of the assembly processes into smaller parts
Confirmation and Demarcation	R6	Occurs intermittently. The first may affirm or disprove users' interpretations or actions. Subsequent ones may break up assembly processes into smaller parts
Clarification and Demarcation	R7	First step is used to reduce confusion and clear doubts. Subsequent ones may break up assembly processes into smaller parts
Refocusing	R8	Redirects user attention to critical aspects of the instruction
Error	R9	A simple user mistake of repeating instructions unnecessarily

R1 repeats are for clarifications that may lead to user self-correction. There are two sub-classes of R1. In R1a, clarification occurs before the user executes the current instruction, leading to self-correction. In R1b, after repeating instruction for clarification, the user makes a task error then self–corrects. For example, the user repeats the instruction, then picks the wrong piece (black instead of yellow) then drops it and picks the correct one (yellow).

R2 types are clarification repeats that reduce confusion and enhance user comprehension. These usually occur after listening to the instruction, in order to clear a user's doubts.

R3 are confirmation repeats that usually occur after listening to an instruction, or carrying out an assembly process. R3s enable users to assess their comprehension and interpretation of agent instruction in comparison to the intended meaning. R3s are also used to assess whether the action taken by the user is correct or incorrect. If these are affirmed, then user self-assurance and confidence levels may improve (R3a). To confirm assistive action taken for example, if the correct piece was picked. R3bs confirms if the correct operative action was done or if the correct assembly procedure was followed (e.g. joining pieces together). R3cs are used for confirmation of assistive action taken, leading to self-correction (e.g. picking pieces, confirming, and then self-correcting. R3ds are used for confirmation of assistive action taken, leading to self-correction of operative action (e.g. picking pieces, confirming, and then assembling. R3es are used for clarification then confirmation of assembly procedures (e.g. joining pieces together).

R4s are multiple repeats that can serve different purposes, but with a focus on executing one assembly step. These types of repeats are used strategically to aid user execution of agent instructions. R4a repeats confirm actions when they occur after the action has taken place, and the user can be seen inspecting the work done while the instruction is been repeated. R4bs are used for clarification, then self-correction, and step-by-step confirmation. R4cs are first used for clarification, followed by self-correction, and finally for confirmation of assembly processes (e.g. joining pieces together). R4ds demarcate and clarify the assembly procedure, refocus a user's attention for easy compliance, and then confirm action taken. R4es demarcate the assembly process, and then confirm the action taken. Finally, R4fs are used to confirm action and self-correct.

R5s are repeats that are used for demarcation. They are used to separate each stage of the assembly process into finely grained parts to simply the assembly process and check information overload.

R6 types are confirmation and demarcation repeats. They occur intermittently between listening and task execution. The first affirms or disproves a users interpretation of agent instruction and its meaning, or the task action taken. The others separate individual stages of the assembly process into smaller parts.

R7 repeats involve clarification and demarcation. The first repeat occurs without any user assembly action, while the subsequent ones occur as needed by the user during assembly or task execution. They tend to reduce or manage user confusion, and separate different stages of the assembly process as the user tries to mitigate the effects of information overload.

R8 repeats refocus and redirect a user’s attention to critical aspects of the instruction. The users repeat the instructions but listens for specific information in the instruction.

R9 repeats occur in error and serve no purpose.

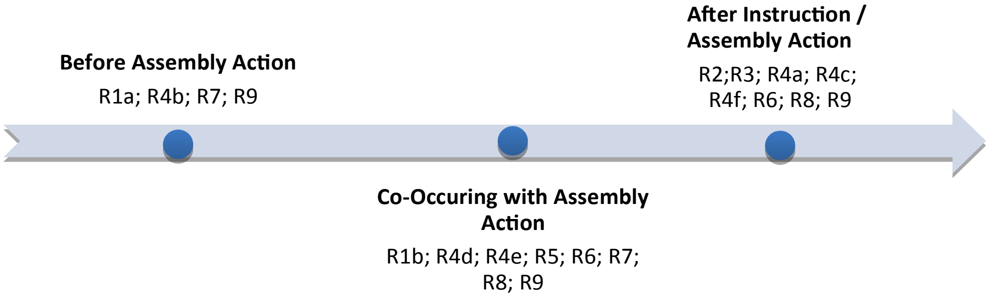

Before Assembly Action
R1a; R4b; R7; R9
After Instruction / Assembly Action
R2;R3; R4a; R4c; R4f; R6; R8; R9
Co-Occuring with Assembly Action
R1b; R4d; R4e; R5; R6; R7; R8; R9

Fig.5. The sequence of instruction repeats.

Repeats are non-arbitrary in nature, which implies that repeats are consistent, reliable and rational communication strategies that promote active listening [Oxford 1993]. Repeats also involve a continuous chain of events. Fig 5 indicates that users ask for repeats before the assembly action or the next instruction occurs such as R1a, R4b, R7, and R9. Repeats can also co-occur with the assembly action in which case they are strategic, for example R1b; R4d; R4e; R5; R6; R7; R8; and R9. Others occur after the assembly action has taken place or the current instruction has been given. These can be confirmatory, for example R2; R3; R4a; R4c; R4f; R6; R8; and R9. However, repeats operate discreetly but can be and are often combined due to timing, and user action. They could also occur at any point in time during the interaction, for example in R9s.

5.4 Assembly Gestures

To address RQ3, the following assembly gestures outlined in this section have emerged (Table 8). These indicate that it is possible for humans to show their intentions to agents in a finely grained nonverbal manner [Kirk et al. 2005]. The function of each gesture is explained in this section using vignettes.

Table 8: Emerging Corpus of Basic Assembly gestures

GESTURAL PHRASE	FUNCTION
Aligning Hand	Enables user to visually and mentally assess the fit of one or more model parts into others
Displaying Hand	Picking and showing model parts for confirmation
Repetitive Opening Hand	Continuously placing the same model parts for initial inspection
Picking Hand	Enables user to select model parts
Joining Hands	Enables User t connect model parts together

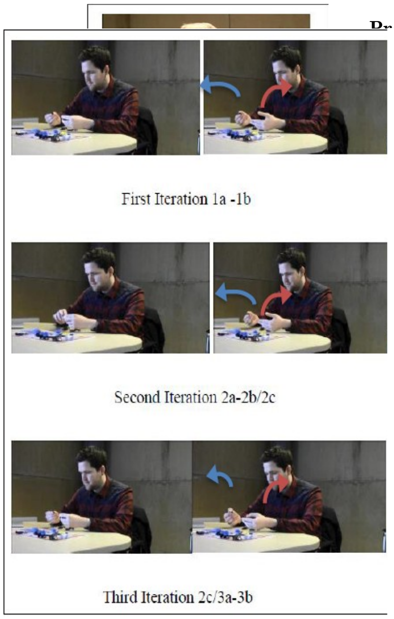
5.4.1 Aligning hands. Aligning hands are similar to “mimicking hands” in an assisted assembly task, discussed in Kirk et al. [2005: 11]. This is because they enable the user to order and discover the fit of the assembly pieces.

Fig. 6. Aligning Hands.

Aligning hands enable the user to visually and mentally assess the fit of the parts of the Lego model. This allows them to discover the compatible sizes. In the first vignette, P1 mentally compares the piece with the larger part of the Lego model it will be assembled into. In the second vignette, he now sizes them up physically. In the third, he starts retracting and this retraction is completed in the fourth vignette. Unlike Kendon’s [2004] gestural phases that have three steps, this one has four steps. The retraction is seen as a two-staged process that starts and culminates with the arms at full rest rather than ending mid-air.

5.4.2 Repetitive opening hands. In assembly tasks, opening hands gestures are similar to beat gestures or rhythmics in their form [Ekman and Friesen 1969a; Ekman and Rosenberg 1997]. This is because they occur with two movements of either up and down, inside-outside, or outside-inside.

However, beat gestures are said to contain no semantic content and are not dynamic [Holler and Wilkin 2011]. To avoid confusion with beat gestures, we have classified the emerging gesture as repetitive opening hands. They are repetitive because they occur in successions called iterations, with one leading seamlessly into the next. The form and function of the repetitive opening hands is shown in the vignette below.



Preparation: P1 Initiates the

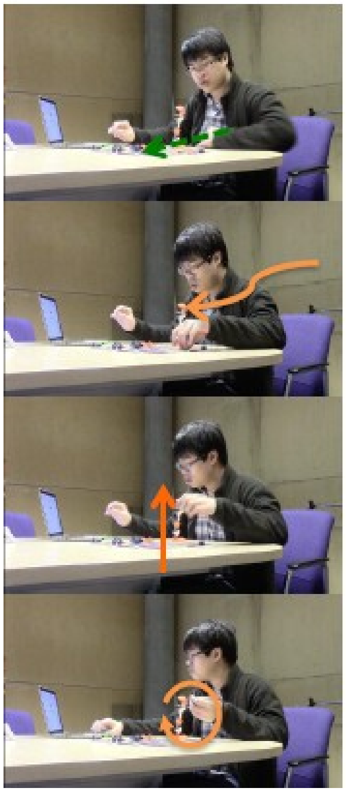
P2 prepares in (1a) by picking the pieces in both hands with fingers curled in a fist. The stroke (S1) occurs in 1b. The hands open sideways/inside out, palm up with digits stretched out placing the pieces picked before P2's eyes for evaluation. P2 repeats the gesture again while listening to the instruction. 2a is preparation while 2b is the stroke (S2) and two handed post-stroke hold (2c). This prepares for the stroke (S3) open sideways/ inside out, palm up with digits stretched out placing the pieces in 3b while listening to the repeated instruction.

Fig. 6. Repetitive opening hands gesture.

In the first iteration vignettes, P2 prepares in 1a and the stroke occurs in 1b. The hands move inside-out, placing the pieces picked before P2's eyes for evaluation. In the second pair of vignettes, P2 repeats the gesture again while listening to the instruction. 2a is preparation, while 2b is the stroke and a post-stroke hold (2c). P2 is still examining the pieces picked for self-evaluation of comprehension levels, and picking the correct pieces signify a successful interpretation of the agent's instruction.

The last pair of vignettes shows that P2 repeats the gesture from the second iteration with a post-stroke hold (2c/3a), which seamlessly prepares for the stroke (3b) while listening to the repeated instruction. Repetitive opening hands enable users to perform task-aiding functions, user self-evaluation of comprehension levels, and the success or otherwise of task progress which may lead to self-correction.

5.4.2 The picking hand. In the assembly process one main hierarchy of tasks is picking kits before putting them together and this is achieved using the picking hand gesture. It is an assistive gesture that leads to others, and may be classified as a lower hierarchy gesture that may determine the success or failure of other aspects of the task. This makes it a crucial gesture in the process of these assemblies. The basic form and function of the gesture is shown in the P12 vignette below.



Preparation: P12 prepares for the gesture with the LH at rest. The broken arrow shows the projected path of the picking hand.

Stroke: P12 performs the stroke by moving the LH, straight towards the piece. The LH is palm down, bending forward with digits held coupler-shaped and the thumb, fore and mid fingers pull up the piece.

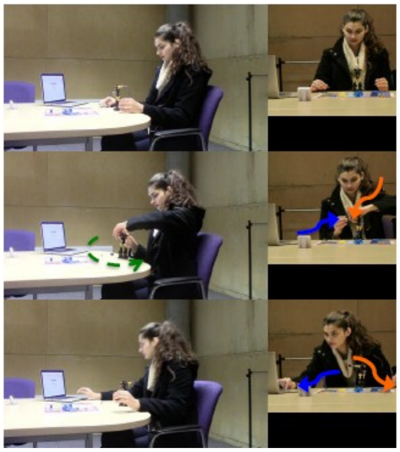
Retraction: P12 begins to retract the LH in preparation for a post-stroke hold.

Post-stroke Hold: P12 retracts the LH turning it over, palm up with the digits still holding the piece to a brief standstill.

Fig.7. The picking hand.

The picking hand gesture in assembly tasks is an iconic representation of the verbs in the instruction. These include *place* and *attach*. This gesture can be performed with one hand, in a back and forth motion, and end with a post stroke hold (Fig 7). It could also be done with two hands, either at once or in quick succession.

5.4.3 The Joining Hand. The joining hands gesture is an operative gesture that enables the user to connect pieces of the Lego model together (Fig 8).



Preparation: P20 initiates the gesture from the RH picking a piece and the LH resting on the table. P20's LH goes up; palm down, with digits held like claws and holds the arm from the top. The RH moves up, palm lateral with sideways grappolo shaped digits brings the piece to join the arm
Stroke: The LH spins (indicated by the broken green arrow) the body around in 180 degrees to position the arm. The RH attaches the piece to the arm and the bottom-up stroke occurs
Retraction: The RH retracts towards the agent while the LH retracts to the table

Figure 8: Joining hands.

The joining hand gesture is an iconic representation of word classes in the instruction [Leff and Sachs 1990]. These include verbs indicating what to do, prepositions indicating where to do it, and adverbs indicating how to do it within the instructions for example. This can be seen in the example instruction below for the model Aquagon:

> *Step 15: Just* ***place*** *the black cylinder into the* ***bottom round hole*** *at the* ***back of*** *the body piece.*

5.5 Emerging Facial Actions

Ekman and Rosenberg [1997] argue that facial actions are categorized as basic emotions, non-basic emotions and emotional attitudes. They discuss that emotions have shared and unique features that differentiate emotions from other affective phenomenon. Three categories of user facial actions emerged from this study so far (Table 9).

Table 9: Formative User Facial Actions

S NO	FACIAL ACTION	ILLUSTRATION
1	Basic Emotions	Surprise, Felt Smile, Neutral
2	Non-basic Emotions	Blend
3	Emotional Attitudes	Frown

Ekman [2007] posits that evolutionary and innate factors, such as muscular and neural activities are responsible for the emotions being characterised as basic, as they evolve to deal with fundamental life tasks such as achievement or failure. Because of this, the users' abilities to execute instructions during the assembly tasks can be observed, in part, through their emotions.

When a combination of basic emotions occurs, this is referred to as a blend or mixed emotional state [Ekman and Rosenberg 1997]. Emotional attitudes are sustained for a longer period, and may involve more than one emotion [Ekman and Friesen 1969a]. Research indicates that some FACs, such as frowns, project emotional attitudes as a salient social signal [Tipples et al. 2002]. The 46 action Units (AU), and action descriptors with underlying facial muscles, described by Ekman and Friesen provide guidelines for analysing FACs and emotions [1969a; 1969b].

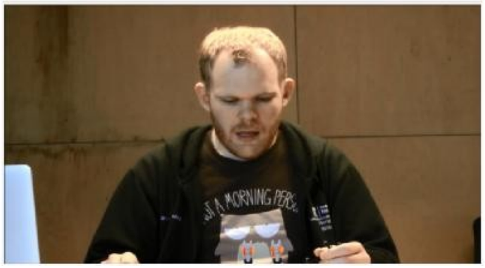


Fig. 9. Slight surprise.

Figure 9 shows P1 expressing a slight surprise (AUs1+2+5B+26). The inner brow is raised a bit, the upper lip is raised and the jaw drops down [Ekman and Rosenberg 1997: 168]. This emotion is triggered by the unexpected voice used by CP to produce the utterance or instruction for the very first time. As well as the facial expression, the participant also displays a fixed attention towards the source of surprise [Cohn et al. 2007]. In this case it the focus is on the model layout where the assembly is executed. The display of surprise may indicate that the participant at that time was unsettled had to repeat the instruction for further understanding and clarification.

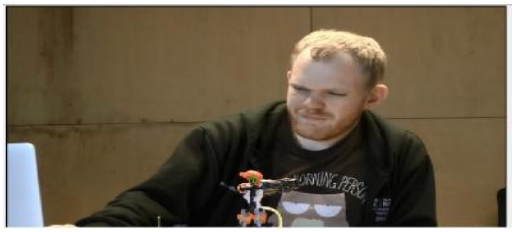


Fig.10. Blend of emotions

A blend of emotions (Fig 10.) is described as a rapid sequence of two basic emotions [Ekman 1992]. This occurs when P1 displays a mixture of enjoyment (AU 6+AU12) and contempt (R12A+R14A) .The cheeks are raised and the lips' corners are pulled back with dimples creating a smug expression [Ekman 2007]. This emotion is targeted towards the participant himself for committing an avoidable error of repeating instructions, and it triggers another emotion shortly after. The participant triggers impatience, though this is not indicated in the picture. The impatience is enacted through a cyclic movement of the participant's hand, which they perform to indicate CP is being to slow in providing the instructions.

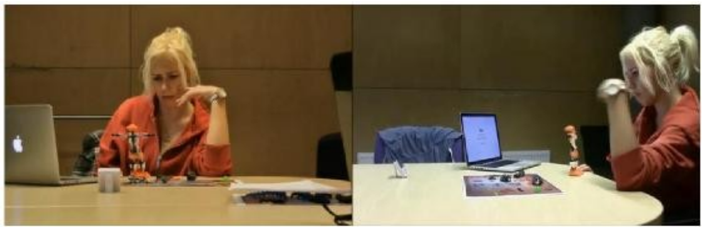


Fig. 11. Frown

P16 frowns (Fig 11) with their eyebrows drawn together and the forehead creased. The forehead gatherer muscle is used to perform this action (AU46+AU64). In a variant of AU4, the corners of the mouth are turned down and the eyes focused downwards [Hinsz and Tomhave 1991]. The fingers on the chin may indicate that P15 is trying to concentrate [Ekman and Rosenberg 1997]. It may also indicate they are paying attention to CP, just as it may indicate a lack of comprehension. In the case of the latter, this is classified as a negative face [Tipples et al. 2002].

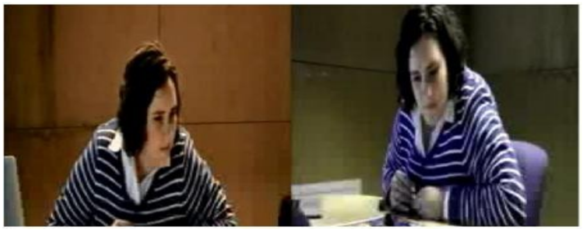


Fig. 12. Static Searching Face

P4 leans forward and listens to the repeated instruction while searching for the piece described, with her eyes sweeping across the table. This is performed using the combination of AUM61, AUM62, AU 57, and AU 23. The static head leans forward (AU57) and combines with eye movements (AUM61+AUM62) [Ekman 1977]. The orbicularis oculi muscle makes a ‘saccade’ eye movement, one of the sub-classes of oculomotor movement [Dell’Osso 1994]. It is used to track an object without head movement. The onset of the symmetrical (AU14) dimpler that aids the lip tightener is immediately preceded or accompanied by eye movement to the left. This is followed to the right, as sweeping eyes indicate that she is searching for the pieces and only stops when she gets the right piece. This facial action is neither a negative or positive face [Tipples et al. 2002]. It may also indicate indifference towards the agent but focus on the task.

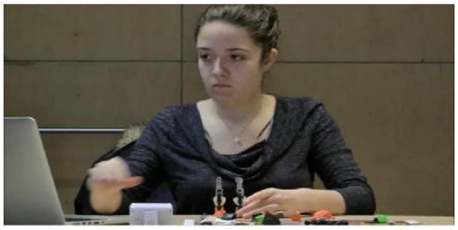


Fig. 13. Neutral Face

P17s neutral face (“AU0” in [Ekman and Rosenberg 1997]) is not showing any emotion as their facial muscles are at rest. This was also a reaction to CP and may indicate that the participant is comfortable the interaction.

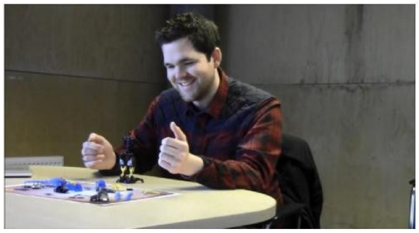


Fig.14. Felt smile

P2 actually experiences a positive emotion (Fig 14) elicited by positive physical, verbal and tactile stimulation [Ekman et al. 1980]. This is performed with the combination several facial muscles: the cheek raiser (AU6), the lip puller (AU12) and the lid tightener (AU7). This smile is a positive face [Tipples et al. 2002]. This is the opposite of a fake smile, and may indicate that the participant is enjoying the interaction and finds VA.

6. DISCUSSION

The first approach described in Section 4 aimed to evaluate the effects of a verbal agent instructor’s voice on the users’ perceptions of vague language. The quantitative measures observed that the human voice VA was perceived as significantly more likeable, less annoying, and more humanlike than the two synthesised voices (CL; CP). VA was also observed to be more coherent than CP. CL and VA were seen to have significant differences in coherence. This may be a result of CP being unable to pronounce some of the items from the VL model as successfully as CL. CL and VA were also seen to equally enable the completion of tasks in comparison to CP.

The use of vague language in VA was seen as more positive than the synthesised voices (4.1.1). Similarly, the general attitudes towards VA were more positive than those towards the two synthesised voices. These attitudes towards the voices support previous research on the preferences for human voices [Cowan et al. 2012; Georgila et al. 2012]. The findings also suggest that, if an agent designer wishes to improve likeability and other positive perceptions of their agent by its users, then using a more humanlike voice may be the ideal option. This may also be true if the aim is to create a younger sounding voice. If these are not of great concern then a synthesised voice may be preferable, if only for logistical and financial benefits.

Similarly, if using vague language, then it appears that using a human voice is the preferable option. However, human likeness in an agent's voice may not always be sufficient for using vague language in agent instructors. Successfully incorporating "humanlike" language in HAI contexts may not solely depend on improving the quality of the voice. Although improvements were seen, there appear to still be barriers to the acceptance of vague agent instructors, regardless of the voice being used. When the vague language did not meet a user's expectations of an agent's speech, disparities between expectations and realities sometimes emerged. While the voice contributed towards this, there were other factors such the agents drawing from both agent and human group identities that had been categorised by previous experiences in the users.

It is advisable then to tread carefully when attempting to use vague language, and possibly similar linguistic features in instruction giving contexts. Outside of this instruction-based context, however, results may differ. Negatively marked perceptions of the agent, even with the voice actor, were also attributed to the context of the interaction – the vague language did not always suit the instruction-based nature of the task. A less controlled environment, or a more leisurely and creative context with a less rigid output, may achieve greater success. Nevertheless, there was still a notable and successful difference between previous studies using vague language with only synthesised voices [Clark et al. 2014].

In analysing the qualitative data, further discussion points arose surrounding agent identity and their place in society. It would appear that prior interactions with agents can have a strong effect on how future ones will transpire, at least when concerning those that instruct, advise or in some other way provide help for their users. Moreover, there was a strong sense of distancing and separation between a human using vague language and an agent using vague language. Even when using the voice actor, some participants still struggled to assign an age to them and sometimes perceived it as just another machine. This raises questions as to features of agent identities, particularly within concepts such as age and the roles agents take on in interaction. Furthermore, there are questions as to what salient linguistic and non-linguistic variables can affect how the users of agents perceive these identities.

The second approach in this paper also raised some interesting discussion points. Repeats maybe very important in instruction giving and by implication advise giving contexts because they involve a systematic process that provides the compass for measuring user feedback during interaction. The user employs repeats to communicate directly with the agent for example, asking for repeats or the next instruction. The study indicates that repeats are used for different purposes in interaction. These include the need for clarification, confirmation of instruction and action taken. Others are used to break up tasks into manageable chunks, redirect user focus, and enhance user self-confidence.

The results indicate that assembly gestures are operative in nature (Table 8) used to explicitly carry out the assembly task set by the agent such as 'aligning hands. They may depict the active or operative verb in the instruction for example, when the instruction says *"Basically, find the end that is a bit more narrow than the other one and just attach the side ball joints to the sockets on the legs"* the user's gesture depicts the acts of picking and measuring the fit of attaching the side ball joints to the socket in the aligning hands gesture.

The other assembly gesture is assistive in nature because they do not in themselves carry out the instructions but they help the user to execute operative gestures. They are lower in hierarchy to operative gestures but are required for operative gestures to succeed. The examples include the repetitive opening hand and the showing hand gestures.

Six types of facial actions were observed in users during interaction. Of the seven samples presented, CP elicited four facial actions (neutral, frown, smug, and slight surprise) representing 66% of the distribution. The frown, smug and slight surprise are negative faces elicited by CP. VA elicited the felt smile – a positive face indicating a positive perception of the agent's identity [Tipples et al. 2002]. CL elicited the static searching face – a neutral face. The faces elicited by VA and CL represented 17% of the distribution respectively. These facial actions indicate that VA may be the most likeable, while CP may be the most annoying because it elicited more negative faces than the rest, at a ratio of 3:1, while CL elicits indifference in interaction.

From the discussion above, the study generally supports the idea that people assign personalities to agents [Lee 2010; Nass et al. 1995]. The use of vague language in human and synthesised voices affects the user's social and behavioural reactions towards the agent during interaction. This could be due to changes that may not conform to how people expect particular agent voices should use vague language.

Despite the small sample size of nonverbal feedback, furthering this work could have some implications for agent designers. The combination of categories of emotions, gestural corpus and taxonomy of repeats could be programmed into agents by designers to create more adaptive advice giving agents that can use these attributes to manage interaction. This is evident in human interaction where emotions are crucial to the development and regulation of interpersonal relationship in any interaction context [Ekman and Rosenberg 1997]. Moreover, emotions provide information to co-interactants about events, responses and probable next behaviour [Ekman and Friesen 1969] this may enhance advice givers' emotive functionality as they may become more aware of human emotions than the present agents. In addition, emotion enables the organism to deal with interpersonal encounters as they emerge and adapt to deal with them in the future and by implication, advice givers may have the ability to predict user intention from records of previous interaction. We believe that just as in human interaction, recognition and prediction skills may enhance advice giving agent's ability to scaffold user interaction experience when difficulties arise.

7. CONCLUSIONS

This paper discusses a two-way approach of evaluating user interaction with vague agent instructors in the context of Lego model assembly tasks. The first approach analyses the effect of human and synthesised voices on user perceptions of agents and their use of vague language. The second focuses on user nonverbal behaviour in interaction with the agent. Findings suggest that there is an argument to be made for using human rather than synthesised voices for agents using vague language. This is supported by the results of the nonverbal feedback. However, the general acceptance of vague language in this context is still fairly low, and this could be a result of the context itself and the position in which humans place agents in society with regards to the linguistic and social characteristics they should be using. The second approach in this paper stimulates discussion on the effects of user nonverbal feedback as a lens for assessing HAI. Future research in this area could focus on expanding these types of approaches to other contexts beyond that of instruction giving. As our interactions with software agents increases rapidly, there is a need to develop new discourse models and frameworks of analysis. Analyses of this kind of data, collected from both laboratory settings and language use in everyday contexts, play a crucial part in furthering our understanding of what it means to interact with software agents.

REFERENCES

Svenja Adolphs, Sarah Atkins, and Kevin Harvey. 2007. Caught Between Professional Requirements and Interpersonal Needs: Vague Language in Healthcare Contexts. *Vague Language Explored* (2007), 62–78. DOI:http://dx.doi.org/10.1057/9780230627420_4

Maneesh Agrawala et al. 2003. Designing effective step-by-step assembly instructions. *ACM SIGGRAPH 2003 Papers on - SIGGRAPH '03* (2003). DOI:http://dx.doi.org/10.1145/1201775.882352

Jens Allwood. 1993. Feedback in second language acquisition. *Adult Language Acquisition* 2 (August 1993), 196–236.

Gisle Andersen. 1998. The Pragmatic Marker like from a Relevance-theoretic Perspective. *Discourse Markers Descriptions and theory Pragmatics & Beyond New Series* (1998), 147. DOI:http://dx.doi.org/10.1075/pbns.57.09and

Oscar Kin-Chung Au, Chiew-Lan Tai, and Hongbo Fu. 2012. Multitouch Gestures for Constrained Transformation of 3D Objects. *Computer Graphics Forum* 31, 2pt3 (2012), 651–660. DOI:http://dx.doi.org/10.1111/j.1467-8659.2012.03044.x

Stephen Axon, Janet Speake, and Kevin Crawford. 2012. ‘At the next junction, turn left’: attitudes towards Sat Nav use. *Area* 44, 2 (2012), 170–177. DOI:http://dx.doi.org/10.1111/j.1475-4762.2012.01086.x

Marian Bartlett et al. 2010. Insights on Spontaneous Facial Expressions from Automatic Expression Measurement. *Dynamic Faces Insights from Experiments and Computation* (2010), 211–238. DOI:http://dx.doi.org/10.7551/mitpress/9780262014533.003.0015

Holly P. Branigan, Martin J. Pickering, Jamie Pearson, and Janet F. Mclean. 2010. Linguistic alignment between people and computers. *Journal of Pragmatics* 42, 9 (2010), 2355–2368. DOI:http://dx.doi.org/10.1016/j.pragma.2009.12.012

Penelope Brown and Stephen C. Levinson. 1987. *Politeness: some universals in language usage*, Cambridge: Cambridge University Press.

Deborah Cameron. 2001. *Working with spoken discourse*, London: SAGE.

Nina C. Capone and Karla K. Mcgregor. 2004. Gesture Development. *J Speech Lang Hear Res Journal of Speech Language and Hearing Research* 47, 1 (January 2004), 173. DOI:http://dx.doi.org/10.1044/1092-4388(2004/015)

Ronald Carter. 2004. *Language and creativity: the art of common talk*, London: Routledge.

Joanna Channell. 1994. *Vague language*, Oxford: Oxford University Press.

Leigh Michael Harry Clark, Khaled Bachour, Abdulmalik Ofemile, Svenja Adolphs, and Tom Rodden. 2014. Potential of imprecision. *Proceedings of the second international conference on Human-agent interaction - HAI '14* (2014). DOI:http://dx.doi.org/10.1145/2658861.2658895

Jeffrey F. Cohn, Zara Ambadar, and Paul Ekman. 2007. Observer-based measurement of facial expression with the Facial Action Coding System. In James A. Coan & John J. B. Allen, eds. *Handbook of emotion elicitation and assessment*. Oxford: Oxford University Press, 203–221.

Malcolm Coulthard. 1992. *Advances in spoken discourse analysis*, London: Routledge.

Benjamin R. Cowan, Holly P. Branigan, and Russell Beale. 26th Annual BCS Interaction Specialist Group Conference on People and Computers. In *Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers*. Swindon: British Computer Society, 39–48.

Joan Cutting. 2007. *Vague language explored*, Basingstoke: Palgrave Macmillan.

Nils Dahlbäck, Qianying Wang, Clifford Nass, and Jenny Alwin. 2007. Similarity is more important than expertise. *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07* (2007). DOI:http://dx.doi.org/10.1145/1240624.1240859

Louis F. Dell'Osso. 1994. Evidence suggesting individual ocular motor control of each eye (muscle). *Journal of Vestibular Research* 45, 5 (1994), 335–345.

G. Donato, M.s. Bartlett, J.c. Hager, P. Ekman, and T.j. Sejnowski. 1999. Classifying facial actions. *IEEE Transactions on Pattern Analysis and Machine Intelligence IEEE Trans. Pattern Anal. Machine Intell.* 21, 10 (1999), 974–989. DOI:http://dx.doi.org/10.1109/34.799905

Dörnyei Zoltán. 2007. *Research methods in applied linguistics: quantitative, qualitative, and mixed methodologies*, Oxford: Oxford University Press.

Paul Ekman and Wallace V. Friesen. 1969a. The Repertoire of Nonverbal Behavior: Categories, Origins, Usage, and Coding. *Semiotica* 1, 1, 49–98.

Paul Ekman and Wallace V. Friesen. 1969b. Nonverbal leakage and clues to deception. *Psychiatry* 32, 1, 88–106.

Paul Ekman. 1977. Biological and Cultural Contributions to Body and Facial Movement. In John Blacking, ed. *he Anthropology of the Body*. London: Academic Press, 34–84.

Paul Ekman, Wallace V. Freisen, and Sonia Ancoli. 1980. Facial signs of emotional experience. *Journal of Personality and Social Psychology* 39, 6 (1980), 1125–1134.

Paul Ekman, Wallace V. Friesen, and Maureen O'sullivan. 1988. Smiles when lying. *Journal of Personality and Social Psychology* 54, 3 (1988), 414–420.

Paul Ekman. 1992. An argument for basic emotions. *Cognition & Emotion PCEM* 6, 3 (January 1992), 169–200. DOI:http://dx.doi.org/10.1080/02699939208411068

Paul Ekman and Erika L. Rosenberg. 1997. *What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)*, New York: Oxford University Press.

Paul Ekman. 2003. *Emotions revealed: recognizing faces and feelings to improve communication and emotional life*, New York: Times Books.

Paul Ekman. 2007. *Emotions revealed: recognizing faces and feelings to improve communication and emotional life*, London: Macmillan.

Dezheng Feng and Kay L. O'Halloran. 2012. Representing emotive meaning in visual images: A social semiotic approach. *Journal of Pragmatics* 44, 14 (2012), 2067–2084. DOI:http://dx.doi.org/10.1016/j.pragma.2012.10.003

Dezheng Feng and Kay L. O'halloran. 2013. The multimodal representation of emotion in film: Integrating cognitive and semiotic approaches. *Semiotica* 2013, 197 (2013). DOI:http://dx.doi.org/10.1515/sem-2013-0082

Suzanne Fleischman and Marina Yaguello. 2004. Discourse markers across languages? Evidence from English and French. *Discourse Across Languages and Cultures Studies in Language Companion Series* (2004), 129–147. DOI:http://dx.doi.org/10.1075/slcs.68.08fle

B.j. Fogg and Clifford Nass. 1997. Silicon sycophants: the effects of computers that flatter. *International Journal of Human-Computer Studies* 46, 5 (1997), 551–561. DOI:http://dx.doi.org/10.1006/ijhc.1996.0104

Bruce Fraser. 1990. Perspectives on politeness. *Journal of Pragmatics* 14, 2 (1990), 219–236. DOI:http://dx.doi.org/10.1016/0378-2166(90)90081-n

Bruce Fraser. 2010. Pragmatic Competence: The Case of Hedging. *New Approaches to Hedging* (2010), 15–34. DOI:http://dx.doi.org/10.1163/9789004253247_003

Kallirroi Georgila, Alan Black, Kenji Sagae, and David R. Traum. 2012. LREC. In Istanbul, 3519–3526.

Erving Goffman. 1959. *The presentation of self in everyday life*, Garden City, NY: Doubleday.

Erving Goffman. 1967. *Interaction ritual; essays in face-to-face behavior*, Chicago: Aldine Pub. Co.

Maartje M.a. De Graaf, Somaya Ben Allouch, and Tineke Klamer. 2015. Sharing a life with Harvey: Exploring the acceptance of and relationship-building with a social robot. *Computers in Human Behavior* 43 (2015), 1–14. DOI:http://dx.doi.org/10.1016/j.chb.2014.10.030

Ioulia Grichkovtsova, Michel Morel, and Anne Lacheret. 2012. The role of voice quality and prosodic contour in affective speech perception. *Speech Communication* 54, 3 (2012), 414–429. DOI:http://dx.doi.org/10.1016/j.specom.2011.10.005

M.a.k. Halliday. Language as system and language as instance: The corpus as a theoretical construct. *Directions in Corpus Linguistics Proceedings of Nobel Symposium 82 Stockholm, 4-8 August 1991*. DOI:http://dx.doi.org/10.1515/9783110867275.61

Ursula Hess and Robert E. Kleck. 1990. Differentiating emotion elicited and deliberate emotional facial expressions. *European Journal of Social Psychology Eur. J. Soc. Psychol.* 20, 5 (1990), 369–385. DOI:http://dx.doi.org/10.1002/ejsp.2420200502

V.B. Hinsz and J.A. Tomhave. 1991. Smile and (Half) the World Smiles with You, Frown and You Frown Alone. *Personality and Social Psychology Bulletin* 17, 5 (January 1991), 586–592. DOI:http://dx.doi.org/10.1177/0146167291175014

Judith Holler and Katie Wilkin. 2011. An experimental investigation of how addressee feedback affects co-speech gestures accompanying speakers' responses. *Journal of Pragmatics* 43, 14 (2011), 3522–3536. DOI:http://dx.doi.org/10.1016/j.pragma.2011.08.002

Andrzej Huczynski. 2004. *Influencing within organizations*, London: Routledge.

Nicholas R. Jennings, Luc Moreau, David Nicholson, Sarvapali Ramchun, Stephen Roberts, Tom Rodden and Alex Rogers. 2014. Human-agent collectives. *Communications of the ACM Commun. ACM* 57, 12 (2014), 80–88. DOI:http://dx.doi.org/10.1145/2629559

Carey Jewitt. 2009. *The Routledge handbook of multimodal analysis*, London: Routledge.

Ing-Marie Jonsson and Nils Dahlbäck. 2011. I Can't Hear You? Drivers Interacting with Male or Female Voices in Native or Non-native Language. *Lecture Notes in Computer Science Universal Access in Human-Computer Interaction. Context Diversity* (2011), 298–305. DOI:http://dx.doi.org/10.1007/978-3-642-21666-4_33

Andreas H. Jucker, Sara W. Smith, and Tanja Lüdge. 2003. Interactive aspects of vagueness in conversation. *Journal of Pragmatics* 35, 12 (2003), 1737–1769. DOI:http://dx.doi.org/10.1016/s0378-2166(02)00188-1

Adamn Kendon. 1980. Gesticulation and speech: Two aspects of the process of utterance. In Mary R. Key, ed. *The Relationship of Verbal and Nonverbal Communication*. 25. The Hague: Mouton Publishers, 207–227.

Adam Kendon. 2004. *Gesture: visible action as utterance*, Cambridge: Cambridge University Press.

Youjin Kim and Nicole Tracy-Ventura. 2013. The role of task repetition in L2 performance development: What needs to be repeated during task-based interaction? *System* 41, 3 (2013), 829–840. DOI:http://dx.doi.org/10.1016/j.system.2013.08.005

David Kirk, Andy Crabtree, and Tom Rodden. Ways of the Hands. *Ecscw 2005*, 1–21. DOI:http://dx.doi.org/10.1007/1-4020-4023-7_1

J. Klein, Y. Moon, and R.w. Picard. 2002. This computer responds to user frustration: *Interacting with Computers* 14, 2 (2002), 119–140. DOI:http://dx.doi.org/10.1016/s0953-5438(01)00053-4

Almut Koester. 2007. 'About Twelve Thousand or So': Vagueness in North American and UK Offices. *Vague Language Explored* (2007), 40–61. DOI:http://dx.doi.org/10.1057/9780230627420_3

George Lakoff. 1973. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *J Philos Logic Journal of Philosophical Logic* 2, 4 (1973). DOI:http://dx.doi.org/10.1007/bf00262952

C.M. Laserna, Y.-T. Seih, and J.W. Pennebaker. 2014. Um . . . Who Like Says You Know: Filler Word Use as a Function of Age, Gender, and Personality. *Journal of Language and Social Psychology* 33, 3 (2014), 328–338. DOI:http://dx.doi.org/10.1177/0261927x14526993

Marianne Latinus and Pascal Belin. 2011. Human voice perception. *Current Biology* 21, 4 (2011). DOI:http://dx.doi.org/10.1016/j.cub.2010.12.033

Eun Ju Lee, Clifford Nass, and Scott Brave. 2000. Can computer-generated speech have gender? *CHI '00 extended abstracts on Human factors in computer systems - CHI '00* (2000). DOI:http://dx.doi.org/10.1145/633458.633461

Eun-Ju Lee. 2010. The more humanlike, the better? How speech type and users' cognitive style affect social responses to computers. *Computers in Human Behavior* 26, 4 (2010), 665–672. DOI:http://dx.doi.org/10.1016/j.chb.2010.01.003

Michael Leff and Andrew Sachs. 1990. Words the most like things: Iconicity and the rhetorical text. *Western Journal of Speech Communication* 54, 3 (1990), 252–273.

Jay L. Lemke. 1990. Technical discourse and Technocratic Ideology. *Learning, Keeping and Using Language Selected papers from the Eighth World Congress of Applied Linguistics, Sydney, 16–21 August 1987. Volume 2* (1990), 435. DOI:http://dx.doi.org/10.1075/z.lkul2.31lem

Andy Lücking, Kirsten Bergman, Florian Hahn, Stefan Kopp, and Hannes Rieser. 2012. Data-based analysis of speech and gesture: the Bielefeld Speech and Gesture Alignment corpus (SaGA) and its applications. *Journal on Multimodal User Interfaces J Multimodal User Interfaces* 7, 1-2 (December 2012), 5–18. DOI:http://dx.doi.org/10.1007/s12193-012-0106-8

Richard E. Mayer, W.Lewis Johnson, Erin Shaw, and Sahiba Sandhu. 2006. Constructing computer-based tutors that are socially sensitive: Politeness in educational software. *International Journal of Human-Computer Studies* 64, 1 (2006), 36–42. DOI:http://dx.doi.org/10.1016/j.ijhcs.2005.07.001

David McNeill. 1992. *Hand and mind: what gestures reveal about thought*, Chicago: University of Chicago Press.

Wendy Mcmillan. 2014. 'They have different information about what is going on': emotion in the transition to university. *Higher Education Research & Development* 33, 6 (June 2014), 1123–1135. DOI:http://dx.doi.org/10.1080/07294360.2014.911250

Albert Mehrabian. 1972. *Nonverbal communication*, Chicago: Aldine-Atherton.

Robert S.K. Miles, Julie Greensmith, Holger Schnadelbach, and Jonathan M. Garibaldi. 2013. Towards a method of identifying the causes of poor user experience on websites. *2013 13th UK Workshop on Computational Intelligence (UKCI)* (2013). DOI:http://dx.doi.org/10.1109/ukci.2013.6651314

Stuart Moran, Nadia Pantidi, Khaled Bachour, Joel E. Fischer, Martin Flintham, Tom Rodden, Simon Evans and Simon Johnson. 2013. Team reactions to voiced agent instructions in a pervasive game. *Proceedings of the 2013 international conference on Intelligent user interfaces - IUI '13* (2013). DOI:http://dx.doi.org/10.1145/2449396.2449445

Clifford Nass, Jonathan Steuer, and Ellen R. Tauber. 1994. Computers are social actors. *Proceedings of the SIGCHI conference on Human factors in computing systems celebrating interdependence - CHI '94* (1994). DOI:http://dx.doi.org/10.1145/191666.191703

Clifford Nass, Youngme Moon, B.j. Fogg, Byron Reeves, and D.Christopher Dryer. 1995. Can computer personalities be human personalities? *International Journal of Human-Computer Studies* 43, 2 (1995), 223–239. DOI:http://dx.doi.org/10.1006/ijhc.1995.1042

Clifford Nass, B.j. Fogg, and Youngme Moon. 1996. Can computers be teammates? *International Journal of Human-Computer Studies* 45, 6 (1996), 669–678. DOI:http://dx.doi.org/10.1006/ijhc.1996.0073

Clifford Nass and Youngme Moon. 2000. Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues J Social Isssues* 56, 1 (2000), 81–103. DOI:http://dx.doi.org/10.1111/0022-4537.00153

Clifford Nass and Kwan Min Lee. 2001. Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied* 7, 3 (2001), 171–181. DOI:http://dx.doi.org/10.1037/1076-898x.7.3.171

Andreea I. Niculescu. 2011. *Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality*, Enschede: University of Twente.

Fatma Demirci Orel and Ali Kara. 2014. Supermarket self-checkout service quality, customer satisfaction, and loyalty: Empirical evidence from an emerging market. *Journal of Retailing and Consumer Services* 21, 2 (2014), 118–129. DOI:http://dx.doi.org/10.1016/j.jretconser.2013.07.002

Rebecca L. Oxford. 1993. Research update on teaching L2 listening. *System* 21, 2 (1993), 205–211. DOI:http://dx.doi.org/10.1016/0346-251x(93)90042-f

M. Pinto, J.A. Cordon, and R.Gomez Diaz. 2010. Thirty years of information literacy (1977--2007): A terminological, conceptual and statistical analysis. *Journal of Librarianship and Information Science* 42, 1 (June 2010), 3–19. DOI:http://dx.doi.org/10.1177/0961000609345091

Ellen F. Prince, Joel Frader, and Charles Bosk. 1982. On hedging in physician-physician discourse. *Linguistics and the Professions* 8 (1982), 83–97.

Byron Reeves and Clifford Nass. 1996. *The media equation: how people treat computers, television, and new media like real people and places*, Stanford, CA: CSLI Publications.

Tobias Rodemann, Martin Heckmann, Claudius Gläser, Frank Joublin, and Christian Goerick. 2010. Towards Speech Acquisition in Natural Interaction on ASIMO. *JRSJ Journal of the Robotics Society of Japan* 28, 1 (2010), 18–22. DOI:http://dx.doi.org/10.7210/jrsj.28.18

Tim Rowland. 2007. 'Well Maybe Not Exactly, but It's Around Fifty Basically?': Vague Language in Mathematics Classrooms. *Vague Language Explored* (2007), 79–96. DOI:http://dx.doi.org/10.1057/9780230627420_5

Emanuel A. Schegloff, Gail Jefferson, and Harvey Sacks. 1977. The Preference for Self-Correction in the Organization of Repair in Conversation. *Language* 53, 2 (1977), 361. DOI:http://dx.doi.org/10.2307/413107

Matthias Scheutz, Rehj Cantrell, and Paul Schermerhorn. 2011. oward Humanlike Task-Based Dialogue Processing for Human Robot Interaction. *AI Magazine* 32, 4 (2011), 77–84.

Megan Strait, Cody Canning, and Matthias Scheutz. 2014. Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance. *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction - HRI '14* (2014). DOI:http://dx.doi.org/10.1145/2559636.2559670

Rie Tamagawa, Catherine I. Watson, I.Han Kuo, Bruce A. Macdonald, and Elizabeth Broadbent. 2011. The Effects of Synthesized Voice Accents on User Perceptions of Robots. *Int J of Soc Robotics International Journal of Social Robotics* 3, 3 (February 2011), 253–262. DOI:http://dx.doi.org/10.1007/s12369-011-0100-4

Jason Tipples, Anthony P. Atkinson, and Andrew W. Young. 2002. The eyebrow frown: A salient social signal. *Emotion* 2, 3 (2002), 288–296. DOI:http://dx.doi.org/10.1037/1528-3542.2.3.288

Cristen Torrey, Susan R. Fussell, and Sara Kiesler. 2013. How a robot should give advice. *2013 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (2013). DOI:http://dx.doi.org/10.1109/hri.2013.6483599

Hugh Trappes-Lomax. 2007. Vague Language as a Means of Self-Protective Avoidance: Tension Management in Conference Talks. *Vague Language Explored* (2007), 117–137. DOI:http://dx.doi.org/10.1057/9780230627420_7

William M.K. Trochim. 2001. *Research methods knowledge base*, Cincinnati, OH: Atomic Dog Pub.

Keiko Tsuchiya. 2013. Listenership Behaviours in Intercultural Encounters. *Pragmatics & Beyond New Series* (August 2013). DOI:http://dx.doi.org/10.1075/pbns.236

Ning Wang, W.Lewis Johnson, Richard E. Mayer, Paola Rizzo, Erin Shaw, and Heather Collins. 2008. The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies* 66, 2 (2008), 98–112. DOI:http://dx.doi.org/10.1016/j.ijhcs.2007.09.003

Ning Wang, W.Lewis Johnson, and Jonathan Gratch. 2010. Facial Expressions and Politeness Effect in Foreign Language Training System. *Intelligent Tutoring Systems Lecture Notes in Computer Science* (2010), 165–173. DOI:http://dx.doi.org/10.1007/978-3-642-13388-6_21

Michael Wooldridge and Nicholas R. Jennings. 1995. Intelligent agents: theory and practice. *The Knowledge Engineering Review Knowl. Eng. Rev.* 10, 02 (1995), 115. DOI:http://dx.doi.org/10.1017/s0269888900008122

Xiangxin Zhu and D. Ramanan. 2012. Face detection, pose estimation, and landmark localization in the wild. *2012 IEEE Conference on Computer Vision and Pattern Recognition* (2012). DOI:http://dx.doi.org/10.1109/cvpr.2012.6248014