Exploring verbal uncanny valley effects with vague language in computer speech

Leigh Clark, Abdulmalik Ofemile & Benjamin R. Cowan

Abstract Interactions with speech interfaces are growing, helped by the advent of intelligent personal assistants like Amazon Alexa and Google Assistant. This software is utilised in hardware such as smart home devices (e.g. Amazon Echo and Google Home), smartphones and vehicles. Given the unprecedented level of spoken interactions with machines, it is important we understand what is considered appropriate, desirable and attractive computer speech. Previous research has suggested that the overuse of humanlike voices in limited-communication devices can induce uncanny valley effects – a perceptual tension arising from mismatched stimuli causing incongruence between users' expectations of a system and its actual capabilities. This chapter explores the possibility of verbal uncanny valley effects in computer speech by utilising the interpersonal linguistic strategies of politeness, relational work, and vague language. This work highlights that using these strategies can create perceptual tension and negative experiences due to the conflicting stimuli of computer speech and 'humanlike' language. This tension can be somewhat moderated with more humanlike than robotic voices, though not alleviated completely. Considerations for the design of computer speech and subsequent future research directions are discussed.

L. Clark
School of Information & Communication Studies, University College Dublin, Ireland
e-mail: leigh.clark@ucd.ie

A. Ofemile
School of English, University of Nottingham, UK
e-mail: abdulmalik.ofemile@nottingham.ac.uk

B.R. Cowan
School of Information & Communication Studies, University College Dublin, Ireland
e-mail: benjamin.cowan@ucd.ie

1 Introduction

As a mode of interaction, speech can affect peoples' perceptions of others in terms of identity, personality, power and attractiveness (Cameron, 2001; Coulthard, 2013; Goffman, 2005; Zuckerman & Driver, 1988). Speech can impact these perceptions in both the language used and the voice quality used to produce it; the latter defined here as "those characteristics which are present more or less all the time that a person is talking" (Abercrombie, 1967, p.91 in Laver, 1980, p.1). As with human-human interaction (HHI), this impact on perceptions can be seen in human-computer interaction (HCI), where speech has become a more prominent mode of interaction. This prominence has been accelerated with the advent of intelligent personal assistants (IPAs) such as Amazon Alexa and Google Assistant featuring in home-based smart speakers like Amazon Echo and Google Home, as well as in mobile devices and vehicles. These are in addition to longer-standing speech-based technologies like interactive voice response (IVR) and navigation systems. Although we are beginning to understand more about how people use and communicate with these types of devices (Cowan et al., 2017; Luger & Sellen, 2016; Porcheron, Fischer, Reeves, & Sharples, 2018; Porcheron, Fischer, & Sharples, 2017), less is known about the psychological and behavioural effects of speech interface design choices on users (Clark, Cabral & Cowan, 2018).

While we are aware that design choices in speech-based HCI can affect user experience (UX) and interaction behaviour, we are still lacking theoretical understandings and subsequent design considerations supporting them (Clark et al., in press). Consequently, it is not always clear what linguistic or voice styles may be appropriate, desirable or even attractive to users in HCI. Mimicking aspects of humanness in speech interfaces, for example, may not always be an appropriate design objective and can result in systems being perceived as creepy or even deceitful (Aylett, Cowan, & Clark, 2019). Recent research (Moore, 2017a) has argued that humanlike voices are not always appropriate for non-human artefacts, as they may heighten peoples' expectations of what artefacts are capable of, in contrast to more robotic voices. This heightened perception of humanness can result in a gap between users' perceptions of a system's abilities or *partner models* and the reality of its limitations observed through interaction (Cowan et al., 2017). As well as the quality of a system's voice, there are also less explored questions as to what are considered appropriate styles of language for computer speech, and how humanlike or 'machinelike' they are expected to be (Clark, 2018; Clark et al., 2019).

This chapter explores the concepts of three interpersonal linguistic strategies – politeness, relational work and vague language (VL) – as a lens to examine the possibility of *verbal uncanny valley effects* that exist in users' perceptions towards both voice and language in computer speech. This may underpin some of the user behaviour and perceptions of appropriateness, desirability and attractiveness directed towards speech interfaces in previous research, as well as peoples' expectations and partner models of their computer interlocutors. It is hoped these

discussions may drive theoretical understandings of our interactions with speech interfaces, which may in turn encourage design considerations in the field.

2 Uncanny Valley

The *uncanny valley* hypothesis suggests that non-human artefacts approaching close to humanlikeness, but retaining obvious differences from human norms, can induce negative responses from people due to one or more obvious differences from expected human appearance or behaviour (Mori, 1970; Mori, MacDorman, & Kageki, 2012). These responses may be referred to as concepts like eeriness, revulsion, or a sense of unease, signifying perceptions of undesirable or unattractive characteristics. Disfluencies between appearance and motion, for instance, may be more disliked than entities displaying more congruent features – contrasting an android that is humanlike in appearance yet displaying robotic movements with an all human and all robot alternative (Carr, Hofree, Sheldon, Saygin, & Winkielman, 2017).

While empirical evidence for the uncanny valley is somewhat scarce, a review of uncanny valley research papers highlighted support for two perceptual mismatch hypotheses (Kätsyri, Förger, Mäkäräinen, & Takala, 2015). The first of these hypotheses suggests that uncanny valley effects arise due to mismatches between the humanlikeness of different sensory cues (e.g. obviously non-human eyes on a fully humanlike face). The second hypothesis posits that the effects occur because of a higher sensitivity towards exaggerated features on more humanlike characters that differ from expected humanlike norms (e.g. “grossly enlarged eyes” (Kätsyri et al., 2015, p.7)). Similar explanations for uncanny valley effects are discussed by Moore (2012). In developing a Bayesian explanation for the uncanny valley effect, Moore points to conflicting cues creating a perceptual distortion and subsequent perceptual tension at category boundaries. These categories refer to stimuli that are discriminately perceived as being different from one another. Stimuli perceived to be at the boundaries of these categories may incur more perceptual distortion than those stimuli perceived to be prototypical examples of those categories.

Whereas most uncanny valley research has focused on the visual, there are an increasing number of works that include audio as an additional modality of interest in exploring perceptual mismatches. Grimshaw (2009) discusses the concept of an audio uncanny valley, with the view that further theoretical understandings may be useful for sound design in horror-based computer games in creating perceptions of fear and apprehension. The author provides examples of features that may induce uncanny valley effects, including uncertainty about the location of sound sources and exaggerated articulation of the mouth whilst speaking. Mitchell et al. (2011) and Meah and Moore (2014) explored the concepts of misaligned voice and face cues (or mismatched stimuli) in robots and humans. Both experiments showed that

mismatches in voice and face (e.g. robotic voice and human face or human voice and robotic face) result in higher ratings of perceived eeriness than matched stimuli.

These experiments give credence to the uncanny valley existing in audio as well as visual stimuli, although the focus in the above work is on multimodal cues and the audio is primarily centred on the voice quality. With the increasing number of speech interfaces, users are exposed to unprecedented levels of primarily speech-based interactions with machines. However, there remain important design considerations on what is considered appropriate speech output by speech interfaces. Moore (2017a), for example, highlights the proliferation of humanlike rather than more robotic sounding voices in computer speech is not always an appropriate design choice. Using humanlike voices can create mismatches between users' expectations of a machine's capabilities and the reality of what it can achieve through speech. This may result in unsuccessful engagement with speech-based, non-human artefacts. Less is understood as to what may be considered appropriate language in spoken interactions with machines – perceptual mismatches may also occur on a linguistic as well as a voice level, potentially resulting in unwanted negative effects to UX (Clark, 2018). The subsequent sections of this chapter reflect on recent research into the use of interpersonal linguistic strategies in spoken computer instructions, and discuss the possible boundaries of appropriate language use (as opposed to solely the appropriate humanlike synthesis choices) in light of uncanny valley theories and mismatched stimuli (Clark, Bachour, Ofemile, Adolphs, & Rodden, 2014; Clark, Ofemile, Adolphs, & Rodden, 2016).

3 Politeness and Relational Work

The concept of politeness is often discussed in terms of Brown and Levinson's (1987) work that associates politeness with the concept of *face* – the social self-image that we present to others during interaction (Goffman, 1955). This self-image is dependent on sociocultural and contextual factors and dynamically progresses between and within interactions. Face theory discusses it being in speakers' own interests to avoid damaging the face of oneself or the face of others during interaction. Conducting this is known as *face-work*.

In Brown and Levinson's (1987) research, face-work can be accomplished using politeness strategies. *Positive face* refers to desires of being liked and approved. Positive politeness strategies include showing group membership between partners, paying attention to the wants and desires of others, and presenting approval. *Negative face* refers mainly to the desire not to be imposed upon by others. Negative politeness strategies often focus on minimising this potential imposition. This can be accomplished by being indirect rather than direct, for example when issuing instructions or making requests that may create an imbalance of power.

Relational work seeks to expand Brown and Levinson's (1987) politeness theory to include the whole polite-impolite spectrum (Locher, 2004; 2006; Locher &

Watts, 2005; 2008). This includes all work by individuals for the "construction, maintenance, reproduction and transformation of interpersonal relationships among those engaged in social practice" (Locher & Watts, 2008, p.96). As with face-work and politeness described above, relational work is similarly discursive and on-going (Locher & Watts, 2005; 2008; Watts, 2003).

3.1 Politeness in Machines

While there are disagreements in politeness and relational work, the politeness strategies discussed in this chapter focus on the polite end of the relational work spectrum and discuss a combination of positive and negative politeness strategies discussed in Brown and Levinson's (1987) theory. In some previous research, politeness strategies have been explored in both the HCI and human-robot interaction (HRI) communities, although the visual modality and/or the use of embodiment were as prominent as speech. For example, Wang et al. (2008) employed politeness strategies in a Wizard-of-Oz experiment providing tutorial feedback to students. The tutorial interface contained visual features – in the form of text and an animated robotic character that produces gestures – and text-to-speech (TTS) synthesis that would appear to come from the robotic character. In comparing polite and direct feedback, the authors note that students receiving the polite tutorial feedback learned better than those receiving the direct feedback. Furthermore, politeness appeared to be especially effective for students who displayed a preference for indirect help or were judged to have less ability to complete the task.

In an HRI-based experiment, positive attitudinal results were observed. Torrey, Fussell and Kiesler (2013) conducted a study in which participants observed videos of human and robot helpers giving advice to a person learning to make cupcakes. In creating the communication conditions, the authors used combinations of hedges and discourse markers. Hedges (e.g. *sort of*, *I guess*) are described by the authors as a negative politeness strategy mitigating the force of messages and reducing threats to a listener's autonomy. The authors acknowledge that descriptions of discourse markers (e.g. *like*, *you know*) have no standard definition[1], though for the purposes of their study they are described in similar terms hedges in being used to "soften commands" (Torrey et al., 2013, p.277). Four communication conditions were created: direct (no hedges/discourse markers); hedges with discourse markers; hedges without discourse markers; discourse markers without hedges. Results of the experiment showed that hedges and discourse markers as individual strategies improved perceptions towards helpers in terms of considerateness, likeability, and

[1] Discourse markers may also be referred to, amongst other terms, as *discourse particles, pragmatic particles* and *pragmatic expressions*. Their purposes can include switching topics, marking boundaries between segments of talk, helping to conduct linguistic repair and being used as hedging devices (Jucker & Ziv, 1998).

the helper being controlling compared to the direct condition. However, the combination of the two strategies did not show significant differences compared to the individual strategies. While positive improvements in perceptions towards both human and robot helpers were observed, participants only observed videos of interactions with helpers, rather than interact with any themselves.

In a similar study, Strait, Canning and Scheutz (2014) analysed both observations and actual interactions with robots providing advice in a drawing task. The authors created an experiment comparing three different interaction modalities: remote third-person (observations of interactions); remote first person (one-to-one with a robot via a laptop); and co-located first person (one-to-one with robot in the same room). As with the experiment by Torrey et al. (2013), two communication conditions were presented. The indirect condition used a combination of positive politeness strategies (e.g. giving praise, being inclusive) and negative politeness strategies (e.g. being indirect, using discourse markers), whereas the direct condition referred to the absence of these strategies in the robot helper's speech. A further condition was included on the robot's appearance, which compared one robot with a more humanlike appearance and another with a more typical robotic appearance. The results of the experiment showed politeness strategies in the indirect speech condition improve ratings of likeability and reduced ratings of perceived aggression when compared to the direct speech condition. Improved ratings for considerateness were also observed in indirect speech, but only in the remote third-person interaction modality. The findings showed that previous results from observations of interaction of robots do not necessarily transfer to actual interaction.

3.2 Politeness in Non-Embodied Computer Speech

The above studies highlight the mixed user responses towards different types of machines and interaction modalities using politeness strategies, focusing in particular on interactions with partners who are embodied or are represented visually. Many modern speech interface technologies like Google Assistant can include a minimal amount of visual output, depending on the device being used but do not necessarily include embodied features.

With this in mind, two further studies explored the use of politeness strategies in HCI, in which participants were tasked with constructing models under the instruction of a speech interface (Clark et al., 2014; Clark et al., 2016). In both studies, VL was used to create indirectness as a form of overall negative politeness strategy[2]. VL refers to language that is deliberately imprecise and can achieve a

[2] These were: adaptors e.g. *more or less*, *somewhat* (reduce assertiveness, minimize imposition); discourse markers e.g. *so, now* (structure talk, mitigate assertive impact of utterance); minimisers e.g. *just, basically* (structure talk, reduce

wide range of functional and interpersonal goals (Channell, 1994). For example, lexical hedges like *just* and *partly* can be used as a tension-management device to playdown the perceived significance of research during academic conferences (Trappes-Lomax, 2007). Furthermore, vague nouns such as *thing* and *whatsit* can be used to replace a typical noun if speakers and listeners have both established what the vague nouns are referring to (Channel, 1994). While not all VL has functions in being polite, this is the primary purpose of which it used in the speech interface studies – the indirectness and imprecision of VL can contribute to lessening the perception of speakers being too authoritative (McCarthy & Carter, 2006) and help create an informal and less direct atmosphere during interaction.

In the first speech interface study using VL, two communication conditions were developed – a vague condition containing politeness strategies and a non-vague condition excluding these politeness strategies (Clark et al., 2014). Participants were tasked with building Lego models under the verbal instructions of a computer interface, the speech of which was produced by the TTS voice Cepstral Lawrence[3]. During this study, participants interacted with an interface on a MacBook Pro 10.2. This was a minimalistic interface using an HTML file linked to a library of pre-recorded speech files. The interface allowed participants to proceed to the next instruction or repeat a current instruction, with the pace being dictated by the participants. Results of this study showed the non-vague interface was rated as significantly more direct and authoritative than the vague interface. However, post-task interviews revealed participants perceived the vague interface to be inappropriate in terms of its language choice. This was partly a result of the quality of the voice. People's expectations of a relatively robotic voice were matched more with the non-vague interface than the vague interface, with the latter discussed as being insincere and its language more appropriately suited to a more natural (i.e. humanlike) sounding voice.

A follow-up experiment explored vague communication conditions across three different voices (Clark et al., 2016). Two of these were TTS synthesised voices – Cepstral Lawrence as per the previous experiment – and CereProc Giles[4]. The third voice was provided by a professional voice actor who was deemed to sound similar in age and accent to the two synthesised voices. Participants followed verbal instructions to build models using two of the three voices in two separate tasks. These tasks used the same style of interface as the first experiment. Results showed the voice actor was perceived as significantly more likeable, more humanlike, and less annoying than the two synthesised voices. Furthermore, it was perceived as more coherent than Giles and both the voice actor and Lawrence were rated as allowing more task completion than Giles. Analysis of post-task interview data also revealed that VL in both synthesised voices were perceived negatively. Participants

---

perceived difficulty, mitigate utterance impact) and vague nouns e.g. *thing, bit* (improve language efficiency) (Clark et al., 2016).

[3] https://www.cepstral.com

[4] https://www.cereproc.com

cited it as inappropriate and often commented on the jarring nature between the quality of the voice and the language being used. However, while the voice actor was seen as a more appropriate fit for VL, results were not wholly convincing. Despite the increased naturalness and humanlikeness, participants still highlighted the disparity between the more machinelike nature of the voice and the humanlike nature of the language. Even with a human voice, there were comments discussing it as 'just a machine' that is not capable of executing VL or politeness strategies, unlike other people, due to their inherent interpersonal and social linguistics purposes. This suggests that the medium of speech delivery, in this case a machine, can also impact on perceptions of appropriateness and attractiveness.

4 Implications for Verbal Uncanny Valley Effects

In terms of what may be considered appropriate computer and human speech, the experiments discussed above raise the possibility of category boundaries existing on a linguistic level – verbal uncanny valley effects. While participants could not always explicitly identify individual lexical items that caused negative reactions towards the interfaces, they were able to identify a general disparity between the language being used and the interface that provided the language. Although this was not the case for all participants, there was a general trend towards describing the vague conditions in both experiments as humanlike language, whereas in Clark et al. (2014), the non-vague condition was cited as being appropriately machinelike.

In the sense of the latter, the use of direct and non-vague language was seen to match people's expectations of appropriate language use with a robotic synthesised voice. This is an example of matched speech-based stimuli, whereby categories of preconceived 'machinelikeness' are aligned. Subsequently, there is little discussion about feelings of the uncanny arising, which are focused more on misaligned stimuli (Mitchell et al., 2011; Moore, 2012a). This also draws similarities with Moore's (2012a) discussion of appropriate voices in non-human artefacts. With non-vague and direct instructions provided by a robotic voice, appropriateness is seemingly determined as it matches people's expectations of what their interaction partner is capable of. These expectations and beliefs of what a communicative partner can produce may be referred to as peoples' partner models (e.g. Cowan, Branigan, Obregón, Bugis, & Beale, 2015). Previous research with infrequent users of IPAs has suggested that speech qualities such as regional accents can signal the communicative attributions people make towards artificial assistants (Cowan et al., 2017). Similarly, this may operate with the quality of a system's voice, the language it uses, and how these two relate to one another. A robotic voice may relate more to signals of using direct than indirect language that is absent in relational work, vague language or politeness strategies. In terms of users' expectations, these linguistic concepts may not be seen as residing in the category of appropriate computer speech.

This can be observed in the vague conditions of the two experiments (Clark et al., 2014; Clark et al., 2016). In the synthesised voices in particular – the combination of a robotic sounding voice with language that is used to undertake social goals – creates a mismatch in stimuli. Subsequently, uncanny valley effects can be observed, especially in participants' descriptions of their interactions with the interfaces. In the second experiment (Clark et al., 2016), however, using a pre-recorded human voice appeared to cause less perceived stimuli mismatch in the vague conditions than the synthesised voices. This may indicate that perceived categories of appropriate computer and human speech can be blurred somewhat with the introduction of more humanlike voices – a human voice can signal a perceptual cue of being capable of producing more humanlike language, even in a computer interface. However, the mismatch is not alleviated completely. Other cues, such as the medium and/or context of interaction (laptop interface providing task-based instructions) may alter what is perceived as appropriate speech even with a human voice.

4.1 Identifying Appropriateness in Computer Speech

Indeed, the combination of socially-driven linguistic cues and computer speech output may create a *habitability gap* (Moore, 2017b), whereby there is a gap between a users' model of a system and the reality of the actual system (Hone & Graham, 2000). Users' models of computer speech may not include the use of interpersonal linguistic strategies and subsequently the presentation of actual computer speech that includes these creates feelings of unease or *perceptual tension* (Moore, 2012).

The mismatching of cues and accompanying perceptual tension in spoken interactions with computers and other machines appears strongly linked to perceptions of what is considered appropriate communication. In addition to a possible habitability gap, it may also be the case that perceived inappropriateness of politeness, relational work or vague language in computer speech is aligned with the socially-driven nature of these concepts. Relational work and politeness strategies, for example, are primarily focused on establishing and maintaining interpersonal relationships with other people (Locher & Watts, 2008; Brown & Levinson, 1987). It is debatable as to what extent this can be accomplished in HCI, how achievable this is as a design goal, and how much users would desire this feature in a speech-based device. The social rules that underpin much HHI do not automatically transfer to HCI and the latter may be markedly diminished in comparison. Moore (2017b, p.8) highlights a similar possible phenomenon – that there may be a "fundamental limit" to the linguistic interactions between humans and machines due to them being "*unequal* partners". The very nature of humans and machines means there are inherent differences in capabilities, and this is likely present in the partner models users create in speech-based HCI. When these partner

models clash with experiences, this may lead to negative user experiences and perceptions of inappropriate, undesirable or unattractive speech interface partners.

The social rules underpinning HCI and HHI also do not automatically align. Relational work and politeness strategies are primarily focused on interpersonal relationships. Brown and Levinson's (1987) theory on politeness in particular is strongly associated with the process of face-work during interaction. However, the maintenance of face during interaction with machines is different than with other people – machines do not have a face as such to protect and, in turn, users do not have another self-image they have to consider during interaction. There may be elements of corporate rather than individual self-images present during interaction, and users can still be imposed upon by machines. However, this remains markedly different from interaction with other people. Indeed, recent research observed that, while descriptions of conversations with people often discuss social and interpersonal wants and needs, interactions with machines are described in very functional and tool-like terms (Clark et al., 2019). This may be due to a lack of familiarity and experience from which to draw upon. However, spoken interactions with machines lack many of the conversational complexities seen in human communication, and are often limited to isolated question-answer pairs (Porcheron et al., 2018).

5 Future Work and Considerations for Computer Speech

This chapter has presented the possible existence of verbal uncanny valley effects – that perceptual tension and negative user experiences and attitudes can emerge in spoken interactions with computers when using linguistic strategies that are inherently social and interpersonal. This effect appears to be intensified with more robotic voices and lessened, though not entirely, with more humanlike voices. This differs from previous discussions of an auditory uncanny valley (e.g. Grimshaw, 2009; Meah & Moore, 2014) in that it focuses on both language as well as voice quality, and the relationship between them. Verbal uncanny valley effects suggest there may be category memberships that exist with styles of language that focus on relational work – i.e. that other people are members of this category whereas computers do not become automatic members by virtue of employing the same strategies. Doing so may create an impression of machines encroaching upon the verbal space of people. This is similar to Moore's (2017b) discussion of there being a fundamental limit to spoken interaction between humans and machines. Moore (2015) mentions that endowing machines with features like humanlike voices can create the mismatched stimuli that lead to perceptual tension, and this may also hold true for certain linguistic styles. With similar considerations, it appears that reducing perceptual tension with verbal uncanny valley effects may depend partly on the relationship between voice and language. If using a very robotic voice, interpersonal linguistic strategies may not be appropriate and may be subsequently undesirable

and unattractive. Conversely, if wanting to employ these strategies, a more humanlike voice would be more appropriate. However, there remains the possibility that no matter what voice is used, certain interpersonal language may be evaluated negatively regardless due to fundamental and embedded differences in user expectation between humans and computers as interlocutors.

It is likely that this is not always the case – this argument stops short of saying all types of interpersonal linguistic strategies are off limits. However, there are design choices around voice and language to consider for computers using speech. There are also other choices to consider. The discussions of politeness strategies and VL in this chapter tend to focus on task-based scenarios in HCI. While this is arguably where most speech-based HCI still currently remains at a linguistic level, it may be the case that instruction-giving or advice-giving computers in task-based scenarios are not appropriate vessels for interpersonal language. If the aim of an interaction between speaking computers and humans is fundamentally an interpersonal one (e.g. social talk (Gilmartin, Cowan, Vogel, & Campbell, 2017) or in healthcare dialogues (Bickmore et al., 2018)), then these linguistic styles may be more appropriate. Similarly, the role in which both computer and human play in any given interaction may also influence evaluations of speech – an instruction-giver may be treated differently to a machine that operates more on a peer-level or as a caregiver, due to varying levels of power and exactly what linguistic possibilities these roles may afford. Similarly, human controlled speech synthesis output, such as the use of a vocal synthesiser to create the ability to speak, may be evaluated differently to speech synthesis output that is controlled by a machine. Furthermore, the direction of interaction may have an effect. Previous experiments often focus on speech output only from a system, whereas two-way dialogue may induce different evaluations. Previous research has shown that politeness can be reciprocated back and forth in an interaction with an in-car help system (Large, Clark, Quandt, Burnett, & Skrypchuk, 2017), though the work does not provide insight into people's actual evaluations of the system.

However, while these ideas are rooted in evidence from previous research, there is still the need to test them further. As noted in Section 2, the evidence for the uncanny valley alone is scarce, with Moore's (2012) Bayesian approach offering a rare quantitative verification of its existence. Future research endeavours can explore the concept of a verbal uncanny valley and its effects further in both quantitative and qualitative means, although any notions of a valley in terms of the shape is arguably less important than the effects cause by underlying concepts of fundamental communicative limits. Comparisons with actual human stimuli as well as computers may also prove beneficial. Indeed, quantifying what constitutes 'humanlike' or 'machinelike' communication is a complex process. Given the increasing prevalence of computer speech, what is perceived as 'machinelike' may well change over the years as familiarity with these devices increases. Longitudinal studies may also uncover further evidence on the effects of prolonged interaction with devices and the extent to which this may affect any verbal uncanny valley effects.

6 Summary and Conclusion

Determining what is considered appropriate speech in HCI remains a challenge. Moore (2017a) offers examples of how to determine appropriateness in the voices of non-human artefacts and avoid uncanny valley effects – robotic rather than humanlike in less sophisticated systems may be better at matching users' expectations of a system with reality. Language use, however, is arguably a more complex affair. This chapter discusses three concepts of interpersonal linguistic strategies (politeness, relational work and VL) to explore what may be considered appropriate language use in speech-based HCI. In linking previous experiments on these strategies with research on the uncanny valley, we find that the social rules that underpin human interaction do not automatically transfer to HCI. The concept of face – the social self-image presented to others – is mostly non-existent on the part of the system during interaction. The need to conduct face-work i.e. protecting this self-image, is then diminished. While users can still be imposed upon by an interface, using strategies like politeness and VL may not always be appropriate and may be undesirable. The combination of computer speech and interpersonal language gives rise to perceptual mismatch at the category boundaries between human and computer speech, creating potential for negative user evaluations of systems. Consequently, this raises the potential of verbal uncanny valley effects, whereby the use of very 'humanlike' language creates feelings of perceptual tension in HCI. While a humanlike voice can act as a moderator for these effects, it does not alleviate perceptual tension completely. Future research should explore the empirical testing of the verbal uncanny valley and its effects, identify what linguistic concepts are seen to reside in the category of appropriate and inappropriate computer speech, and understand what further phenomena (like voice) may influence its evaluation by users.

Acknowledgements

This research was funded by a New Horizons grant from the Irish Research Council entitled "The COG-SIS Project: Cognitive effects of Speech Interface Synthesis" (Grant R17339).

References

Abercrombie, D. (1967). Elements of general phonetics (Vol. 203). Edinburgh: Edinburgh University Press.

Aylett, M. P., Cowan, B. R., & Clark, L. (2019). Siri, Echo and Performance: You have to Suffer Darling. In *Proceedings of the 2019 CHI Conference on HumanFactors in Computing Systems*. ACM.

Bickmore, T. W., Trinh, H., Olafsson, S., O'Leary, T. K., Asadi, R., Rickles, N. M., & Cruz, R. (2018). Patient and Consumer Safety Risks When Using Conversational Assistants for Medical Information: An Observational Study of Siri, Alexa, and Google Assistant. *Journal of Medical Internet Research*, *20*(9). https://doi.org/10.2196/11510.

Brown, P., & Levinson, S. C. (1987). *Politeness: Some Universals in Language Usage*. Cambridge University Press.

Cameron, D. (2001). *Working with Spoken Discourse*. SAGE.

Carr, E. W., Hofree, G., Sheldon, K., Saygin, A. P., & Winkielman, P. (2017). Is that a human? Categorization (dis)fluency drives evaluations of agents ambiguous on human-likeness. *Journal of Experimental Psychology: Human Perception and Performance*, *43*(4), 651–666. https://doi.org/10.1037/xhp0000304.

Channell, J. (1994). *Vague Language*. Oxford University Press.

Clark, L. (2018). Social Boundaries of Appropriate Speech in HCI: A Politeness Perspective. *Proceedings of British HCI*.

Clark, L., Cabral, J. & Cowan, B. R. (2018). The CogSIS Project: Examining the Cognitive Effects of Speech Interface Synthesis. In *Proceedings of British HCI*.

Clark, L., Doyle, P., Garaialde, D., Gilmartin, E., Schlögl, S., Edlund, J., … Cowan, B. (2018). The State of Speech in HCI: Trends, Themes and Challenges. *ArXiv Preprint ArXiv:1810.06828*.

Clark, L. M. H., Bachour, K., Ofemile, A., Adolphs, S., & Rodden, T. (2014). Potential of imprecision: exploring vague language in agent instructors (pp. 339–344). ACM Press. https://doi.org/10.1145/2658861.2658895

Clark, L., Ofemile, A., Adolphs, S., & Rodden, T. (2016). A Multimodal Approach to Assessing User Experiences with Agent Helpers. *ACM Trans. Interact. Intell. Syst.*, *6*(4), 29:1–29:31. https://doi.org/10.1145/2983926

Clark, L., Pantidi, N., Cooney, O., Doyle, P., Garaialde, D., Edwards, J., … others. (2019). What Makes a Good Conversation? Challenges in Designing Truly Conversational Agents. *ArXiv Preprint ArXiv:1901.06525*.

Coulthard, M. (2013). *Advances in Spoken Discourse Analysis*. Routledge.

Cowan, B. R., Branigan, H. P., Obregón, M., Bugis, E., & Beale, R. (2015). Voice anthropomorphism, interlocutor modelling and alignment effects on syntactic choices in human−computer dialogue. *International Journal of Human-Computer Studies*, *83*, 27–42. https://doi.org/10.1016/j.ijhcs.2015.05.008

Cowan, B. R., Pantidi, N., Coyle, D., Morrissey, K., Clarke, P., Al-Shehri, S., … Bandeira, N. (2017). 'What can i help you with?': infrequent users' experiences of intelligent personal assistants (pp. 1–12). ACM Press. https://doi.org/10.1145/3098279.3098539

Gilmartin, E., Cowan, B. R., Vogel, C., & Campbell, N. (2017). Exploring Multiparty Casual Talk for Social Human-Machine Dialogue. In *International Conference on Speech and Computer* (pp. 370–378). Springer.

Goffman, E. (1955). On Face-Work. *Psychiatry*, *18*(3), 213–231. https://doi.org/10.1080/00332747.1955.11023008

Goffman, E. (2005). *Interaction Ritual: Essays in Face to Face Behavior*. AldineTransaction.

Grimshaw, M. (2009). The audio Uncanny Valley: Sound, fear and the horror game. *Audio Mostly*, 21–26.

Hone, K. S., & Graham, R. (2000). Towards a tool for the Subjective Assessment of Speech System Interfaces (SASSI). *Natural Language Engineering*, *6*(3–4), 287–303.

Jucker, A. H., & Ziv, Y. (1998). *Discourse Markers: Descriptions and theory*. John Benjamins Publishing.

Kätsyri, J., Förger, K., Mäkäräinen, M., & Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Frontiers in Psychology*, *6*. https://doi.org/10.3389/fpsyg.2015.00390

Large, D. R., Clark, L., Quandt, A., Burnett, G., & Skrypchuk, L. (2017). Steering the conversation: A linguistic exploration of natural language interactions with a digital assistant during simulated driving. *Applied Ergonomics*, *63*, 53–61. https://doi.org/10.1016/j.apergo.2017.04.003.

Laver, J. (1980). The phonetic description of voice quality: Cambridge Studies in Linguistics. Cambridge University Press, Cambridge.

Locher, M. A. (2004). *Power and Politeness in Action: Disagreements in Oral Communication*. Walter de Gruyter.

Locher, M. A. (2006). *Polite behavior within relational work: The discursive approach to politeness*. Walter de Gruyter.

Locher, M. A., & Watts, R. J. (2005). Politeness Theory and Relational Work. *Journal of Politeness Research. Language, Behaviour, Culture*, *1*(1). https://doi.org/10.1515/jplr.2005.1.1.9

Locher, M. A., & Watts, R. J. (2008). *Relational work and impoliteness: Negotiating norms of linguistic behaviour*. Mouton de Gruyter.

Luger, E., & Sellen, A. (2016). 'Like Having a Really Bad PA': The Gulf Between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (pp. 5286–5297). New York, NY, USA: ACM. https://doi.org/10.1145/2858036.2858288

McCarthy, M., & Carter, R. (2006). This that and the other: Multi-word clusters in spoken English as visible patterns of interaction. *Explorations in Corpus Linguistics*, 7.

Meah, L. F. S., & Moore, R. K. (2014). The Uncanny Valley: A Focus on Misaligned Cues. In M. Beetz, B. Johnston, & M.-A. Williams (Eds.), *Social Robotics* (pp. 256–265). Springer International Publishing.

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., & MacDorman, K. F. (2011). A Mismatch in the Human Realism of Face and Voice

Produces an Uncanny Valley. *I-Perception*, *2*(1), 10–12. https://doi.org/10.1068/i0415

Moore, R. K. (2012). A Bayesian explanation of the 'Uncanny Valley' effect and related psychological phenomena. *Scientific Reports*, *2*(1). https://doi.org/10.1038/srep00864

Moore, R. K. (2015). From Talking and Listening Robots to Intelligent Communicative Machines. In *Robots that talk and listen*. de Gruyter.

Moore, R. K. (2017a). Appropriate Voices for Artefacts: Some Key Insights. 1st International Workshop on Vocal Interactivity in-and-between Humans, Animals and Robots.

Moore, R. K. (2017b). Is Spoken Language All-or-Nothing? Implications for Future Speech-Based Human-Machine Interaction. In *Dialogues with Social Robots* (pp. 281–291). Springer, Singapore. https://doi.org/10.1007/978-981-10-2585-3_22

Mori, M. (1970). The uncanny valley. *Energy*, *7*(4), 33–35.

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [from the field]. *IEEE Robotics & Automation Magazine*, *19*(2), 98–100.

Porcheron, M., Fischer, J. E., Reeves, S., & Sharples, S. (2018). Voice Interfaces in Everyday Life. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (p. 640). ACM.

Porcheron, M., Fischer, J. E., & Sharples, S. (2017). 'Do Animals Have A ccents?': Talking with Agents in Multi-Party Conversation (pp. 207–219). ACM Press. https://doi.org/10.1145/2998181.2998298

Strait, M., Canning, C., & Scheutz, M. (2014). Let me tell you! investigating the effects of robot communication strategies in advice-giving situations based on robot appearance, interaction modality and distance (pp. 479–486). ACM Press. https://doi.org/10.1145/2559636.2559670

Torrey, C., Fussell, S. R., & Kiesler, S. (2013). How a robot should give advice (pp. 275–282). IEEE. https://doi.org/10.1109/HRI.2013.6483599

Trappes-Lomax, H. (2007). Vague language as a means of self-protective avoidance: Tension management in conference talks. In *Vague language explored* (pp. 117–137). Springer.

Wang, N., Johnson, W. L., Mayer, R. E., Rizzo, P., Shaw, E., & Collins, H. (2008). The politeness effect: Pedagogical agents and learning outcomes. *International Journal of Human-Computer Studies*, *66*(2), 98–112. https://doi.org/10.1016/j.ijhcs.2007.09.003

Watts, R. J. (2003). *Politeness*. Cambridge University Press.

Zuckerman, M., & Driver, R. E. (1988). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, *13*(2), 67–82. https://doi.org/10.1007/BF00990791