Chapter 1
Trust in vocal human-robot interaction: implications for robot voice design

Ilaria Torre and Laurence White

Abstract Trust is fundamental for successful human interactions. As robots become increasingly active in human society, it is essential to determine what characteristics of robots inuence trust in human-robot interaction, in order to design robots with which people feel comfortable interacting. Many interactions are vocal by nature, yet the vocal correlates of trust behaviours have received relatively little attention to date. Here we examine the existing evidence about dimensions of vocal variation that inuence trust: voice naturalness, gender, accent, prosody and interaction context. Furthermore, we argue that robot voices should be designed with specic robot appearance, function, and task performance in mind, to avoid inducing unrealistic expectations of robot performance in human users.

1.1 Introduction

Trust is an essential foundation for human societies. Numerous approaches have been taken towards understanding the means by which it is negotiated. For background, the reader is referred to texts in biology (Bateson, 2000), evolutionary theory (Harcourt, 1991), sociology (Luhmann, 1979), economics (Berg et al., 1995), and neuroscience (Bzdok et al., 2011). Here, it will suffice to say that trust relates both to attribution – when someone makes a decision to trust someone else – and to states and traits, when someone acts, in the short term or over the long term, in a trustworthy manner.

Human social evolution has made us very sensitive to cues that may provide information about state or trait trustworthiness in others (e.g. Jones and George, 1998),

Ilaria Torre
Department of Electronic and Electrical Engineering, Trinity College Dublin e-mail: torrei@tcd.ie

Laurence White
School of Education, Communication and Language Sciences, Newcastle University e-mail: laurence.white@ncl.ac.uk

to the point that a short extract of someone's speech (McAleer et al., 2014), or a short exposure to someone's face (Willis and Todorov, 2006) are enough to make us form a consistent impression of that person's trustworthiness . As robots increasingly become part of our daily lives, it is important to understand what makes people trust robots and, conversely, how we can design robots to appear trustworthy, in order to facilitate human-robot interaction (HRI) and collaboration. While much effort is put into designing robots to look trustworthy and appropriate for their task (e.g. Saldien et al., 2008; DiSalvo et al., 2002; Lütkebohle et al., 2010; Oh et al., 2006), less consideration is given to designing voices for these robots (e.g. Sandygulova and O'Hare, 2015). As we argue in this chapter, voice is a very powerful cue used in judgements of trustworthiness, and it should be carefully considered – in conjunction with the robot's appearance/ function and with the users' expectations – when designing a robot.

With regard to vocal attractiveness more generally, this chapter considers how characteristics of a robot's voice that contribute to an impression of trustworthiness reinforce, and are reinforced by, features of vocal attractiveness. Dion et al. (1972) used the phrase "what is beautiful is good" to refer to the fact that attractiveness is strongly perceived as related to other positive traits, including that of trustworthiness. Indeed, attractiveness and trustworthiness loaded on the same factor in McAleer et al. (2014)'s study of vocal features. Additional evidence of the close link between attractiveness and trustworthiness comes from neuroscience (Bzdok et al., 2011) and neurobiology (Theodoridou et al., 2009): Bzdok et al. (2011), for example, concluded that specific brain regions, such as the amygdala, might selectively reinforce sensory information with high social importance, such as information concerning potential relationships (e.g.: "Is this person attractive? I might date them in the future."; "Is this person trustworthy? I might collaborate with them in the future."). Here, we examine what characteristics of a robot's voice, by analogy with human voices, contribute to an impression of trustworthiness as a socially-relevant cue for human-robot interaction.

1.2 Trust in voices

Most human communication is predicated on some degree of mutual trust between interlocutors. When we ask a stranger for directions, we trust that they will give us the correct information, to the best of their knowledge (cf, Cooperative Principle, Grice, 1975). Moreover, like the fabled 'boy who cried wolf', untrustworthy communicators tend to be downgraded as interlocutors once their deceitfulness has been exposed.

As the spoken channel is typically our main mode of communication, we have naturally developed vocal means to signal our trustworthiness and to detect it in others. Indeed, the natural tendency to trust speech is mediated by heuristics that give us indicators about when the speaker might not be trustworthy. Not being able to determine a speaker's background can contribute to this impression, as can vocal identifiers of social affiliations disfavoured by the perceiver, or prosodic indicators of aggression or dominance. Conversely, positive evidence for trustworthiness can be inferred from many vocal features, such as accent (e.g. Lev-Ari and Keysar, 2010), prosodic cues (e.g. Miller et al., 1976), or emotional expressions (e.g. Schug et al., 2010).

Regarding accents, the literature suggests that foreign accents tend to be trusted less than native accents (Lev-Ari and Keysar, 2010), and that, within a language, ‘prestigious’ and ‘standard‘ accents are trusted more than regional accents (Giles, 1970). For example, in the context of the UK, Standard Southern British English (SSBE) is generally evaluated as more trustworthy than, for example, typical London or Birmingham accents (Bishop et al., 2005). Furthermore, experimental evidence suggests that such first impressions of trustworthiness might persist over time, despite being mediated by experience of a speaker’s actual behaviour (Torre et al., 2016).

Results are less conclusive, indeed sometimes contradictory, regarding the direction of influence of various prosodic features on trust attributions. For example, O’Connor and Barclay (2017) found that people had greater trust in higher-pitched male and female voices (based on average fundamental frequency, f0). By contrast, Villar et al. (2013), amongst other studies, have found that participants raise their vocal pitch when lying, and Apple et al. (1979) showed that speakers with a high f0 and slow speech rate were rated as ‘less truthful’. A fast speaking rate was found to be a feature of charismatic and persuasive speakers (Jiang and Pell, 2017; Chaiken, 1979), but has also been found to detract from charisma in speech (Niebuhr et al., 2016). Finally, higher pitch and slower speech rate predicted greater trusting behaviour in an economic game (Torre et al., 2016). Such variable results might be due to the fact that the studies employed different methods, such as questionnaires or behavioural measures, and looked at different aspects of trust, such as deception, economic trust, voting behaviour, charisma, and so on. They might also reflect quantitative variation in the prosodic features examined in the different studies: it is unlikely that the relationships between trust attributions and, for example, speech rate or pitch range are strictly linear. Additionally, rather than individual vocal characteristics, it is more likely to be a combination of features that determine the perceiver’s assessment of trustworthiness, along with how vocal features interact with physical appearance, interaction context, and the perceiver’s emotional state.

Voice is a powerful medium through which a diversity of speaker-specific indexical information is transmitted and interpreted, and robot voices are likely to be subject to similar appraisals. Thus, the design of robot voices should be influenced by the nature of the attributions appropriate to the purposes of particular human-robot interactions.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

1.3 Trust in robot voices

People tend to attribute personality traits to computers and robots as if they were human agents (Nass and Lee, 2001; Nass et al., 1995; Walters et al., 2008), and to respond to robots as if they had a personality (Lee et al., 2006). Given also that people attribute traits to human speakers based on subtle speech characteristics (e.g. Torre et al., 2016; McAleer et al., 2014), there is reason to assume that voice information will be used to attribute traits – e.g. of trustworthiness – to robots as well. Thus, voice selection should be an integral part of the overall robot's design. Issues to take into consideration are numerous and diverse, the following being just a selection. Should large robots have lower-pitched voices than small robots, congruent with anthropomorphic expectations about larger larynxes? Should human-like robots have natural human voices? Should robot voices have regional accents? If so, should these be chosen to reflect the accent of the person with whom they are interacting or, for example, to reflect a stereotyped association between particular voice styles and the specific functions that the robot will perform? The latter approach risks reinforcing stereotypes, but ignoring any considerations of voice-function congruency could be problematic for the naturalness of the interaction.

It seems, however, that relatively little attention is currently paid to how the selection of robot voices in HRI research might affect our interaction with robots. For example, McGinn and Torre (2019) conducted an informal survey of researchers whose paper at the Human-Robot Interaction 2018 conference featured a speaking robot. Specifically, they asked if they chose the voice of their robot and, if so, why. Of the 18 responses received, six had used the Nao robot built-in voice, seven had used a voice generated with a Text-To-Speech system, either because it was freely available or because it was the voice that the robot came with, three pre-recorded the voice using actors, and two simply described what the voice sounded like (e.g. "androgynous, child-like voice"). In addition, six of these authors specified that they had adjusted the robot voice in terms of pitch or speech rate to increase intelligibility or to elicit the perception of a particular voice age. Only one author mentioned the accent that the voice had, and only one author mentioned looking for a voice that would specifically suit the task that the robot had to carry out in the experiment. About the reasons for their choice, two authors specified that "it was the only good one" and "because it was open source". Whilst 11 mentioned the gender of the robot's voice, only a minority considered other voice characteristics such as prosody or accent, or the context in which the interaction would take place. However, as we show in this chapter, all of these features influence human perception of robots, and should not be neglected.

Studies experimentally manipulating a robot voice in order to measure its effect on users' perceptions and behaviours are relatively scarce, but here we review work in which a robot's voice was manipulated, or where vocal characteristics were considered in the analyses . As trust is related to other positive traits – a typical 'halo effect' (Dion et al., 1972) – and as studies examining the effect of robot voices on trust are limited, we evaluate voice-based research in human-robot interaction in general, considering the implications for trustworthiness in particular.

1.3.1 Voice naturalness

One key aspect of voice that is often taken into account when designing robots is naturalness. While current efforts in the speech technology community are dedicated to creating the most natural-sounding artificial voices, it might not be the case that people actually prefer interacting with a robot or other artificial agent with a perfect natural-sounding voice (Hinds et al., 2004). For example, Sims et al. (2009) showed that being able to speak with either a synthetic or a natural voice was enough for a robot to be treated as a competent agent: people gave more commands to a robot that had a voice, whether synthetic or natural, and fewer to a robot that communicated with beeps. They hypothesised that people assumed that speechless robots would not understand language, and thus they did not speak either. Within the speaking robot condition, however, participants gave more commands to the synthetic-voiced robot than the natural one: Sims et al. suggested that participants might have thought that a robot with a human voice was more competent and therefore needed fewer commands. Taking a different perspective, Mitchell et al. (2011) argued that incongruence in the human likeness of a character's face and voice can elicit feelings of eeriness. In contrast, Tamagawa et al. (2011) argued that, for the sake of clarity and familiarity, people would prefer such an 'incongruent' robot. In Eyssel et al. (2012), participants were shown a video of a Flobi robot saying: "it's quarter past three" and were asked to rate the robot in terms of anthropomorphism, likeability, psychological closeness, and intentions. The robot had either a natural or a synthetic voice. Interestingly, voice had an effect only on participants' ratings of likeability, with people rating the natural voice higher. On the other hand, in Torre et al. (2018), people implicitly trusted robots with synthetic voices more than those with natural voices when they were behaving trustworthily, but found the opposite effect when the robots were behaving untrustworthily. This also points to the importance of interaction context for robot voice design (1.3.5).

More generally, Hegel (2012) did not find strong evidence that the human likeness of a robot's appearance influenced the perception of its social capabilities. If the same were true for the human likeness of robot *voices*, this would argue that voice naturalness might not be critical for creating feelings of trust. However, another factor to take into account when considering naturalness is listening effort: thus, listening to synthetic voices can increase cognitive load relative to natural voices (Simantiraki et al., 2018; Francis and Nusbaum, 2009). In turn, high cognitive load hinders strategic thinking and can lead to trust misplacement, for example to trusting untrustworthy individuals (Duffy and Smith, 2014; Samson and Kostyszyn, 2015). This suggests that – especially if the robot is meant to sustain an extended vocal communication with a person – it should be given a natural – or high-quality synthetic, – voice, notwithstanding any contradictions with the robot's mechanical looks.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

1.3.2 Voice gender

Talking specifically about trust, research on human-human interaction has not found consistent differences in trust judgments towards men or women (e.g. Nass and Brave, 2005; Chaudhuri and Gangadharan, 2007; Boenin and Serra, 2009; Slonim and Guillen, 2010). Given that people's mental models of humanoid social robots are generally similar to human models (Lee et al., 2005; Kiesler and Goetz, 2002), it would be reasonable to expect a lack of overall difference it might be that intrinsic differences in the trusting individuals will remain], when it comes to trusting a 'female' or 'male' robot. Indeed, Crowell et al. (2009) failed to find any difference in how people reacted to a mechanical robot that had either a female or a male voice. Thus, in terms of voice design, the straightforward expectation would be that robots that are designed to look more feminine or masculine should have a voice corresponding to their apparent gender.

The problem of voice gender selection may be further simplified by the fact that many robots are not perceived as having a clearly defined gender. For example, in a recent study (partially described in Torre et al., 2018), we used a Nao robot with two different natural female voices, with participants interacting with both. At the end of the experiment, a random sample of the 120 participants was asked what gender they thought the robots had. Of the 66 randomly sampled participants, 23 said they thought the robot was always female, 17 always male, 20 did not associate any gender, and 6 associated a different gender to the two robots they played with. This suggests that even a natural female voice does not consistently convey information about the gender of the robot with that voice. Similarly, the majority of participants in Walters et al. (2008) who interacted with a robot that had either a pre-recorded male voice, a pre-recorded female voice, or a synthesised voice, gave either a male or a neutral name to the robot, even when the robot had a female voice.

Thus, it seems that the gender of a robot voice does not necessarily influence whether people will perceive the robot to have the same gender. However, describing a study involving 9-11-year-old children, Sandygulova and O'Hare (2015) suggested that children assigned a gender to a Nao robot on the basis of the voice alone. This was a synthetic male or female voice. However, participants heard all the possible voices in succession with the same robot, and so a contrast effect may have contributed to the gender attribution being based on voice in this case.

While there is no evidence that voice gender influences a positive human-robot interaction, it is possible that it could interact with presumed gender-specific knowledge (e.g. Powers et al., 2005). As discussed later (section 1.3.5), the context in which the interaction takes place might be more important for trustworthy voice design than voice gender as an isolated feature.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

1.3.3 Voice accent

Everyone has an accent. The term 'accent' refers to systematic patterns of realisation of the sounds of a language – phonetic and phonological – that people belonging to certain geographically- or socially-defined groups tend to have in common (Lippi-Green, 1997). Accents thus provide immediate information about whether or not two interlocutors belong to similar social and/or regional groups, information that we tend to implicitly use in judgements of trustworthiness (e.g. Kinzler et al., 2009). Specifically, in-group membership elicits favourable first impressions, including with robots (Kuchenbrandt et al., 2013). Given that every speaker has an accent, and that these accents affect the way we interpret interpersonal communication, should speaking machines have purpose-specific human accents?

There is, unsurprisingly, evidence of straightforward accent preferences in interactions with robots. For example, children based in Ireland showed a preference towards male and female UK English over US English in a Nao robot (Sandygulova and O'Hare, 2015). We can also contribute some survey data regarding overall preferences for robot accents. All the participants of various UK-based studies run over three years were asked what accent they would like a robot to have. The question was open-ended, so we re-coded the answers to fit in broad categories (e.g. 'West Country' and 'South West' would both be coded as 'South West'; labels such as 'English', 'British', 'RP' would be coded as 'SSBE'). Figure 1.1 shows these standardised answers from all 503 participants who answered this question. As the figure shows, the majority of respondents answered with 'SSBE', followed by 'Neutral' accent (which in the UK is also likely to mean the non-regional SSBE), followed by 'Irish'. All of the respondents were native British English speakers, with the following self-reported regional identities: southwest England (58%), southeast England (22%), Midlands (8%), Wales (5%), East Anglia (3%), with participants from northeast England, northwest England, and Scotland comprising almost all of the remaining 3-4%. As shown in Figure 1.1, very few people reported a preference for a robot to have a machine-like voice. There were also relatively few preferences for a regional accent reflecting one's own origins: 58% of respondents were from the southwest but only about 5% of all respondents said they would like the robot to have a southwestern accent (which here we use to encompass Bristol, Cornwall, Devon, Plymouth or general South-West).

Preferences for robot accents may well also be influenced by the nature of the interaction, however. For example, research from Andrist et al. (2015) on the Arabic language showed an interaction between accent and behavior in human-robot interaction (see section 1.3.5): participants believed that robots with the same regional accent as theirs were more credible – when the robots were knowledgeable – than those with a standard accent, whereas robots with standard accents were perceived to be the more credible when the robots had little knowledge. Similar interactions between accents and behaviour are, of course, likely with other languages. For example, Tamagawa et al. (2011) ran two experiments comparing synthesised British, American, and New Zealand English accents. In the first experiment, participants from New Zealand explicitly rated the disembodied UK accent more positively than the US one, while their own New Zealand accent was not rated significantly differently to either of the other accents. In the second experiment, participants were told by a healthcare robot, in one of the three accents, how to take blood pressure measurements on themselves. After the interaction, they completed a questionnaire, and reported more positive emotions towards the New Zealand-accented robot than the US-accented robot, and thought that the New Zealand robot performed better (no other pairwise comparison was statistically significant). These results also point to the differential effect that accents might have in different interaction contexts (e.g. disembodied voice vs. speaking robot).

1.3.4 Voice prosody

Pitch gives information about a speaker's size, being inversely proportional to the body mass (Ohala, 1983). Thus, it might be straightforward to think that bigger robots should have lower-pitched voices than smaller robots. However, as we saw in the discussion on voice gender (section 1.3.2), intuitive assumptions regarding appropriate voices do not necessarily apply in practice, and experimental work is required. In Niculescu et al. (2011), after interacting with a robot with either a high-pitched or a low-pitched female voice, participants' questionnaire responses indicated an overall preference for the higher-pitched voice. In another study, Yilmazyildiz et al. (2012) asked participants which of two voices, with lower or higher pitch, was more suitable for a NAO – a child-like humanoid robot – or a Probo – a green furry elephant-like robot: participants preferred the higher-pitched voice for NAO and the lower-pitched voice for Probo.

Vocal prosody is also a feature that often manifests convergence. Linguistic convergence – sometimes also called adaptation, entrainment, or synchrony, although we prefer the specificity of "convergence" – is a phenomenon by which two speakers tend to unconsciously imitate each other's speech characteristics as interactions proceed (Beňuš, 2014). According to Communication Accommodation Theory (Giles et al., 1991), convergence is a signal of openness and positive attitude – including trust – towards the interlocutor. For example, Manson et al. (2013) showed that people who converged in terms of their speech rate trusted each other more in a Prisoner's Dilemma task. Looking at linguistic convergence more generally, Scissors et al. (2009) found that some types of linguistic similarity positively correlate with behavioural trust in text-based interaction, while others negatively correlate with it: for example, trusting individuals exhibited convergence in the use of words linked to positive emotions, while deceiving individuals exhibited convergence in the use of negative emotions words.

The well-documented occurrence of convergence phenomenon in human-human interaction led researchers to examine it in human-agent interaction as well. In a computer game where participants followed the advice of an owl-shaped avatar that was either converging or diverging from the participants' own prosody, Beňuš et al. (2018) found that female participants followed the advice of the diverging avatar

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

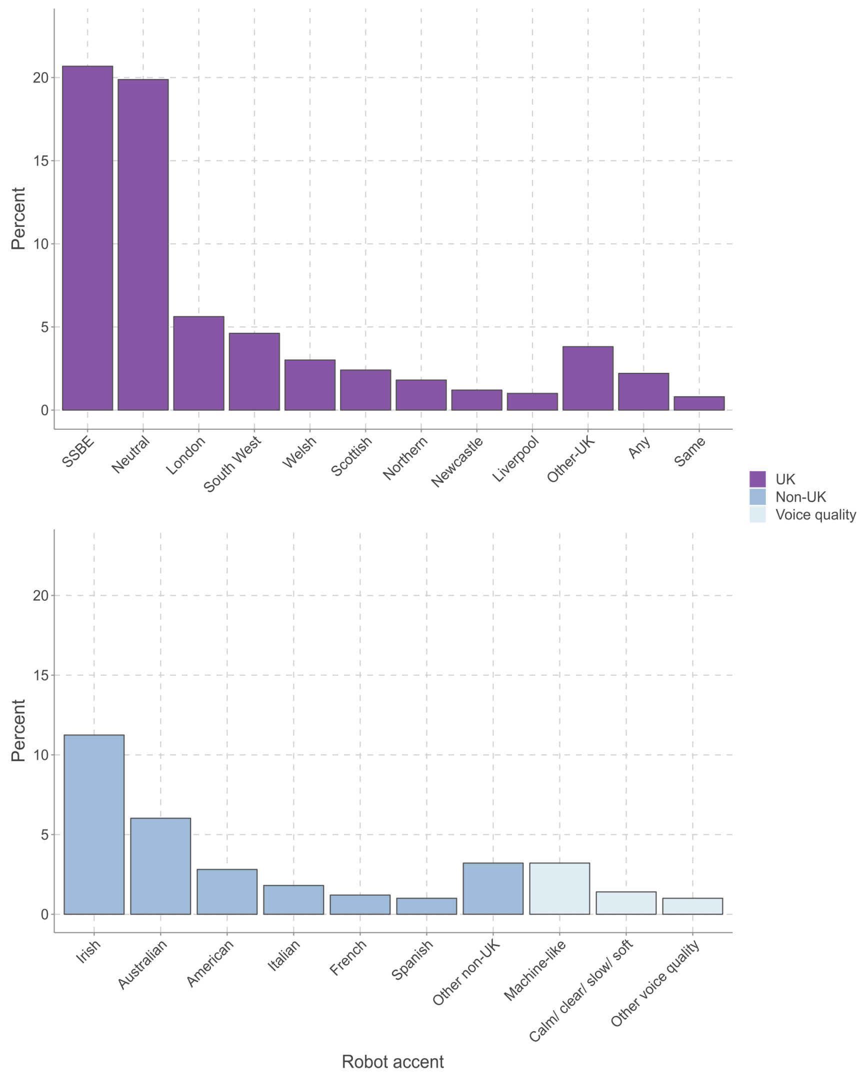

Fig. 1.1 Preference for a robot's accent from a survey of 503 native British English speakers. The question allowed a free response and the bars indicate the overall proportion of all responses that fitted within each accent category (see text for how accent responses were categorised). "Any" means "any accent"; "Same" means "the same accent as me".

more often than the converging one, while there was no effect for male participants. Also contrary to some expectations, Strupka et al. (2016) found that participants tended to diverge prosodically from Keepon robots whose prosody was manipulated. On the other hand, Sadoughi et al. (2017) found that children who played a game with an converging social robot had higher levels of engagement at the end of the interactions than children who played the game with a non-converging robot. The apparent differences in convergence behavior might be due to several factors, notably whether one is concerned with factors promoting convergence by human speakers towards the vocal features of agents or with the impact of convergence by agents on human behaviours and attitudes. Additionally, there may be an influence of age differences in participants, adults in Beňuš et al. (2018) and Strupka et al. (2016), compared with children in Sadoughi et al. (2017): potentially, for example, children may have fewer implicit socio-cognitive biases towards artificial agents. More generally, discrepancies between studies may arise because of intrinsic differences in the artificial voices used. Human speech has been shown to converge with artificial voices in terms of phonetics and prosody when the artificial voice is of high quality, but less so when it is of low quality (Gessinger et al., 2017, 2018). Differences could also be due to appearance contrasts between artificial agents: for example, in the studies reported above, there was an owl-shaped avatar in Beňuš et al. (2018), a small, rudimentarily humanoid robot in Strupka et al. (2016), and a life-size humanoid robot in Sadoughi et al. (2017). Interactions between robot appearance and convergence behaviours cannot be ruled out.

Prosody conveys important information on the emotional state of the speaker (e.g. Bänziger and Scherer, 2005; Aubergé and Cathiard, 2003). In this regard, it is known that displaying a positive emotion generally leads to attributions of other positive traits – including trustworthiness – a typical 'halo' effect (Lau, 1982; Penton-Voak et al., 2006; Schug et al., 2010). Indeed, voice-based Embodied Conversational Agents that were smiling were trusted more than those with a neutral facial expression (Elkins and Derrick, 2013). Smiling in the face also led to trusting avatars and robots more (Krumhuber et al., 2007; Mathur and Reichling, 2016). Thus, a robot expressing positive affect in its prosody could similarly increase the human user's feeling that it can be trusted. The situation-congruent expression of affect might increase trust even when it is not displaying a positive emotion. For example, portraying stress and urgency through the voice increased performance in a joint human-robot collaborative task (Scheutz et al., 2006).

Apart from signalling a speaker's mood or emotional state, prosodic cues also contribute to an individual's vocal profile, that is, what makes a voice unique. Arguably, distinct-looking robots should have different-sounding voices, in order to: a) contribute to the impression that they are individual agents; b) be congruent with their physical appearance; c) elicit personality attributions congruent with the primary functions. In a recent study (partially described in Torre et al., 2018), people played a trust game with robots having different voices. We obtained a natural recording of two female SSBE speakers, which we then resynthesised to sound robotic, thus generating four voices altogether: Speaker 1 natural, Speaker 1 synthetic, Speaker 2 natural, Speaker 2 synthetic. As mentioned earlier (Section 1.3.1), much of the variance in trust was explained by the voice naturalness variable: specifically, people trusted robots with synthetic voices more than those with natural voices when they were behaving trustworthily, but the opposite when the robots were behaving untrustworthily (Torre et al., 2018). However, people also demonstrated greater implicit trust to one of the two speaker voices over the other, both in natural-natural and synthetic-synthetic comparisons. This is consistent with previous studies showing that very fine speech characteristics, which are independent from higher-level features such as accent, affect impression formation (e.g. Gobl and Ní Chasaide, 2003; Trouvain et al., 2006). It also suggests that people's preference for certain individual voices might apply when these voices are embodied in a robot. Thus, idiolectal characteristics, such as those conveyed by prosody, seem to contribute to trusting behaviours as well.

Overall, it seems simplistic to relate trustworthiness judgments purely to isolated vocal features – such as gender, naturalness or pitch – and a holistic view of voice might be better suited for promoting positive interactions, rather than only considering specific individual vocal features.

1.3.5 Voice context and expectations

As discussed earlier, some studies have shown that people perceive robots differently depending on the context in which the interaction takesook place (Sims et al., 2009; Andrist et al., 2015). Thus, the nature of the specific human-robot interaction may affect the optimal characteristics of the robot (see also Torre et al., 2018). For example, Wang et al. (2013) found that, in a favourable context, such as a satisfactory customer/employee call centre interaction, customers with an American English background tended to suppress their negative prejudices towards employees with an Indian English accent. On the other hand, when the interaction was not satisfactory, customers tended not to suppress their accent prejudice (Wang et al., 2013). Similarly, Bresnahan et al. (2002) examined accent perception as a function of the message that the accented speaker was delivering. They recorded two non-native speakers of American English, one very intelligible and one not very intelligible, and one native speaker, reading passages in a 'friend' and 'teaching assistant' condition. Participants were undergraduate students of various ethnic origin, but mostly white Americans. They found that the 'friend' context was judged as more attractive and dynamic than the 'teaching assistant' context, in all accent conditions. Also, participants with and a strong ethnic identity regarded the native accent as higher in status, dynamism and attractiveness, while the opposite was found for participants with a weak ethnic identity, who attributed higher status and attractiveness to the not very intelligent foreign accent, as compared to the native one. Thus, not only the interaction context, but also the specific background context of the human interlocutor is likely to influence the interaction success.

In HRI, Salem et al. (2013) found that participants' perception – in terms of politeness, competency, extraversion, perceived warmth and shared reality – of a receptionist robot differed according to the context of the interaction, which was either goal-oriented or open-ended. By contrast, the variation in the robot's politeness level did not influence participants' perception. Additionally, in the aforementioned study by Sims et al. (2009), participants watched videos of a robot in different scenarios (robot damaged, robot in danger, robot requiring more information, robot has located target, robot has completed task). They found that, for example, participants gave more commands to the robot in the videos where the robot needed assistance, and concluded that a robot's voice should be chosen based on task context. In particular, this would allow for the transmission of pragmatic information which may increase the operation success. For example, in a search and rescue operation, a synthetic voice for a robot might be the appropriate choice, because – whilst it conveys to the person being rescued that the robot is able to help and may be capable of understanding human speech – the fact that the robot voice is not fully human-like could suggest to its human teammates that their input in the operation is still necessary.

As reviewed above, a robot's voice, along with its appearance, will have an influence on the first impressions of that robot's trustworthiness. Given the role of interaction context, however, these first impressions should be validated over long-term interactions with that robot. In fact, several experiments on trusting behaviour in human-machine interaction showed that incongruency between first impressions of trustworthiness and experience of a speaker's actual trustworthiness can drastically reduce trust (Torre et al., 2018). Thus, if a robot's voice gives the impression that the robot will function well, people might have more negative reactions in the case that the robot's performance does not live up to expectations. If it is expected that a robot will operate with some degree of error, perhaps its design (appearance, voice) should reflect the fact that its performance will not always be flawless, so as not to set the users' expectations too high from the beginning (van den Brule et al., 2014). For example, Hegel (2012) found that people attributed higher social capabilities, including honesty, to robots that looked more sophisticated. Whether robots can deliver on their promise of sophisticated performance is a different matter, however, and over-reliance on a robot according to positive first impressions could have major negative consequences (Robinette et al., 2016; Hancock et al., 2011; Salem et al., 2015).

Emotional expression might also elicit different trusting behaviour depending on the interaction context. Van Kleef et al. (2010), in the 'Emotions as Social Information' (EASI) model, suggest that emotions are used to make sense of ambiguous situations, and that their effect depends on the situation in which the interaction takes place, being specifically mediated by its cooperative or competitive nature. Thus, displaying a positive emotion, such as happiness, in a cooperative context will reinforce the parties' belief that everyone is gaining, and will elicit more cooperative behaviours. By contrast, displaying a negative emotion, such as anger, in a cooperative context will hinder future cooperative behaviours. Accordingly, Antos et al. (2011) found that, in a negotiation game, participants tended to select as partners those computer agents which displayed emotions congruent with their actions. Those agents were also perceived as more trustworthy than agents whose emotional expression and action strategy did not match, even though the strategy itself was the same. In summary, emotional expression is helpful only if it is congruent with behaviour.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

1.4 Conclusion

This chapter offers an overview of some of the aspects to consider when designing a trustworthy voice to be used in human-robot interaction. Given that many studies in HRI employing a speaking robot have not carefully considered their robot's voice, the present work aims to be a starting point for subsequent research involving speaking robots.

In particular, we summarised work on the effect that voice naturalness, gender, accent, and prosody can have on trust attributions in human-robot interactions, along with the interactions of such vocal features with the characteristics and demands of the specific human-robot encounter. Naturalness, accent, and prosody seem to be the features with the highest likelihood of shaping trusting behaviour, while voice gender appears secondary. Moreover, carefully controlling for context might be more important than, for example, manipulations of naturalness in the voice: specifically, successful interactions over time may be hindered by inaccurate user expectations arising from mismatches between robot voice features and robot's competence and performance.

It is possible that voice has been a secondary concern in human-robot interaction research so far because vocal interactions have often been scripted, or generated by an imperfect dialogue system, meaning that other aspects of the interaction, such as the robot's movements or attention, might have been prioritised. However, recent advances in the field of natural language and speech processing (such as WaveNet) mean that fluent autonomous human-robot conversations are getting closer to being commonplace. It is time to consider more carefully what the robot's input into these conversations should actually sound like.

Acknowledgements The first author is funded by the European Unions Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 713567. She is also funded by the ADAPT Centre for Digital Content Technology, which is funded under the SFI Research Centres Programme (Grant 13/RC/2016) and is co-funded by the European Regional Development Fund. We are grateful to all the HRI conference authors who generously took time to reply to the survey questions.

References

Sean Andrist, Micheline Ziadee, Halim Boukaram, Bilge Mutlu, and Majd Sakr. Effects of culture on the credibility of robot speech. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI'15*, pages 157–164. ACM, ACM Press, 2015. doi: 10.1145/2696454.2696464.

Dimitrios Antos, Celso De Melo, Jonathan Gratch, and Barbara J. Grosz. The influence of emotion expression on perceptions of trustworthiness in negotiation. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*, 2011.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

William Apple, Lynn A. Streeter, and Robert M. Krauss. Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, 37 (5):715–727, 1979. doi: 10.1037/0022-3514.37.5.715.

Véronique Aubergé and Marie Cathiard. Can we hear the prosody of smile? *Speech Communication*, 40(1-2):87–97, 2003. doi: 10.1016/s0167-6393(02)00077-8.

Tanja Bänziger and Klaus R. Scherer. The role of intonation in emotional expressions. *Speech Communication*, 46(3-4):252–267, 2005. doi: 10.1016/j.specom.2005.02.016.

Patrick Bateson. The biological evolution of cooperation and trust. In Diego Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 14–30. Department of Sociology, University of Oxford, 2000.

Joyce Berg, John Dickhaut, and Kevin McCabe. Trust, reciprocity, and social history. *Games and Economic Behavior*, 10(1):122–142, 1995.

Štefan Beňuš. Social aspects of entrainment in spoken interaction. *Cognitive Computation*, 6(4):802–813, 2014.

Štefan Beňuš, Marian Trnka, Eduard Kuric, Lukáš Marták, Agustín Gravano, Julia Hirschberg, and Rivka Levitan. Prosodic entrainment and trust in human-computer interaction. In *Proceedings of the ninth International Conference on Speech Prosody 2018*. ISCA, 2018. doi: 10.21437/SpeechProsody.2018-45.

Hywel Bishop, Nikolas Coupland, and Peter Garrett. Conceptual accent evaluation: Thirty years of accent prejudice in the UK. *Acta Linguistica Hafniensia*, 37(1): 131–154, 2005. doi: 10.1080/03740463.2005.10416087.

Aurlie Boenin and Daniel Serra. Gender pairing bias in trustworthiness. *The Journal of Socio-Economics*, 38:779–789, 2009.

Mary Jiang Bresnahan, Rie Ohashi, Reiko Nebashi, Wen Ying Liu, and Sachiyo Morinaga Shearman. Attitudinal and affective response toward accented English. *Language & Communication*, 22(2):171–185, 2002.

Danilo Bzdok, Robert Langner, Svenja Caspers, Florian Kurth, Ute Habel, Karl Zilles, Angela R. Laird, and Simon B. Eickhoff. Ale meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure and Function*, 215(3-4):209–223, 2011.

Shelly Chaiken. Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, 37(8):1387, 1979.

Ananish Chaudhuri and Lata Gangadharan. An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, pages 959–985, 2007.

Charles R. Crowell, Matthias Scheutz, Paul Schermerhorn, and Michael Villano. Gendered voice and robot entities: perceptions and reactions of male and female subjects. In *Intelligent Robots and Systems, 2009. IROS 2009. IEEE/RSJ International Conference on*, pages 3735–3741. IEEE, 2009.

Karen Dion, Ellen Berscheid, and Elaine Walster. What is beautiful is good. *Journal of Personality and Social Psychology*, 24(3):285, 1972.

Carl F. DiSalvo, Francine Gemperle, Jodi Forlizzi, and Sara Kiesler. All robots are not created equal: the design and perception of humanoid robot heads. In *Proceedings of the 4th Conference on Designing Interactive Systems: Processes, Practices, Methods, and Techniques*, pages 321–326. ACM, 2002.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

Sean Duffy and John Smith. Cognitive load in the multi-player prisoner's dilemma game: Are there brains in games? *Journal of Behavioral and Experimental Economics*, 51:47–56, 2014.

Aaron C. Elkins and Douglas C. Derrick. The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation*, 22(5):897–913, 2013.

Friederike Eyssel, Dieta Kuchenbrandt, Frank Hegel, and Laura de Ruiter. Activating elicited agent knowledge: How robot and user features shape the perception of social robots. In *Robot and Human Interactive Communication, 2012. RO-MAN 2012. IEEE International Workshop on*, pages 851–857. IEEE, 2012. doi: 10.1109/roman.2012.6343858.

Alexander L. Francis and Howard C. Nusbaum. Effects of intelligibility on working memory demand for speech perception. *Attention, Perception, & Psychophysics*, 71(6):1360–1374, 2009.

Iona Gessinger, Eran Raveh, Sébastien Le Maguer, Bernd Möbius, and Ingmar Steiner. Shadowing synthesized speech – segmental analysis of phonetic convergence. In *Proceedings of Interspeech 2017*, pages 3797–3801, 2017.

Iona Gessinger, Antje Schweitzer, Bistra Andreeva, Eran Raveh, Bernd Möbius, and Ingmar Steiner. Convergence of Pitch Accents in a Shadowing Task. In *Proceedings of the 9th International Conference on Speech Prosody 2018*, pages 225–229, 2018.

Howard Giles. Evaluative reactions to accents. *Educational review*, 22(3):211–227, 1970.

Howard Giles, Nikolas Coupland, and Justine Coupland. Accommodation theory: Communication, context, and consequence. In Howard Giles, Nikolas Coupland, and Justine Coupland, editors, *Contexts of accommodation: Developments in applied sociolinguistics*, pages 1–68. Cambridge University Press, 1991.

Christer Gobl and Ailbhe Ní Chasaide. The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, 40:189–212, 2003.

Herbert P. Grice. Logic and conversation. In Peter Cole and Jerry L. Morgan, editors, *Syntax and semantics: Speech acts*, volume 3, pages 41–58. New York: Academic Press, 1975.

Peter A. Hancock, Deborah R. Billings, Kristin E. Schaefer, Jessie Y. C. Chen, Ewart J. De Visser, and Raja Parasuraman. A meta-analysis of factors affecting trust in human-robot interaction. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 53(5):517–527, 2011.

Alexander H. Harcourt. Help, cooperation and trust in animals. In R. A. Hinde and J. Groebel, editors, *Cooperation and Prosocial Behaviour*, pages 15–26. Cambridge University Press, 1991.

Frank Hegel. Effects of a robot's aesthetic design on the attribution of social capabilities. In *IEEE RO-MAN: The 21st IEEE International Symposium on Robot and Human Interactive Communication*, pages 469–475. IEEE, 2012.

Pamela J. Hinds, Teresa L. Roberts, and Hank Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19(1):151–181, 2004.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

Xiaoming Jiang and Marc D. Pell. The sound of confidence and doubt. *Speech Communication*, 88:106–126, 2017.

Gareth R. Jones and Jennifer M. George. The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, 23(3):531–546, 1998.

Sara Kiesler and Jennifer Goetz. Mental models of robotic assistants. In *Proceedings of the CHI 2002 Conference on Human Factors in Computer Systems*. New York: ACM, 2002.

Katherine D. Kinzler, Kristin Shutts, Jasmine M. DeJesus, and Elizabeth S. Spelke. Accent trumps race in guiding childrens social preferences. *Social Cognition*, 27 (4):623, 2009.

Eva Krumhuber, Antony S. R. Manstead, Darren Cosker, Dave Marshall, Paul L. Rosin, and Arvid Kappas. Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, 7(4):730–735, 2007.

Dieta Kuchenbrandt, Friederike Eyssel, Simon Bobinger, and Maria Neufeld. When a robots group membership matters. *International Journal of Social Robotics*, 5 (3):409–417, 2013.

Sing Lau. The effect of smiling on person perception. *The Journal of Social Psychology*, 117(1):63–67, 1982. doi: 10.1080/00224545.1982.9713408.

Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. Can robots manifest personality? An empirical test of personality recognition, social responses, and social presence in human–robot interaction. *Journal of Communication*, 56(4):754–772, 2006.

Sau-lai Lee, Ivy Yee-man Lau, Sara Kiesler, and Chi-Yue Chiu. Human mental models of humanoid robots. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 2767–2772. IEEE, 2005. doi: 10.1109/robot.2005.1570532.

Shiri Lev-Ari and Boaz Keysar. Why don't we believe non-native speakers? the influence of accent on credibility. *Journal of Experimental Social Psychology*, 46 (6):1093–1096, 2010. doi: 10.1016/j.jesp.2010.05.025.

Rosina Lippi-Green. *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press, 1997.

Niklas Luhmann. *Trust and Power*. John Wiley & Sons Inc, 1979. ISBN 9780471997580.

Ingo Lütkebohle, Frank Hegel, Simon Schulz, Matthias Hackel, Britta Wrede, Sven Wachsmuth, and Gerhard Sagerer. The Bielefeld Anthropomorphic Robot Head "Flobi". In *Robotics and Automation (ICRA), 2010 IEEE International Conference on*, pages 3384–3391. IEEE, 2010.

Joseph H. Manson, Gregory A. Bryant, Matthew M. Gervais, and Michelle A. Kline. Convergence of speech rate in conversation predicts cooperation. *Evolution and Human Behavior*, 34(6):419–426, 2013.

Maya B. Mathur and David B. Reichling. Navigating a social world with robot partners: A quantitative cartography of the uncanny valley. *Cognition*, 146:22–32, 2016.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

Phil McAleer, Alexander Todorov, and Pascal Belin. How do you say Hello? Personality impressions from brief novel voices. *PLoS ONE*, 9(3):e90779, 2014.

Conor McGinn and Ilaria Torre. Can you tell the robot by the voice? An exploratory study on the role of voice in the perception of robots. In *Proceedings of the 14th Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI'19*. ACM, 2019.

Norman Miller, Geoffrey Maruyama, Rex J. Beaber, and Keith Valone. Speed of speech and persuasion. *Journal of Personality and Social Psychology*, 34(4):615, 1976.

Wade J. Mitchell, Kevin A. Szerszen, Amy Shirong Lu, Paul W. Schermerhorn, Matthias Scheutz, and Karl F. MacDorman. A mismatch in the human realism of face and voice produces an uncanny valley. *i-Perception*, 2(1):10–12, 2011.

Clifford I. Nass and Scott Brave. *Wired for speech: How voice activates and advances the human-computer relationship.* Cambridge, MA: MIT Press, 2005.

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

Clifford I. Nass, Youngme Moon, B. J. Fogg, Byron Reeves, and Chris Dryer. Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, 43(2):223–239, 1995.

Andreea Niculescu, Betsy van Dijk, Anton Nijholt, and Swee Lan See. The influence of voice pitch on the evaluation of a social robot receptionist. In *2011 International Conference on User Science and Engineering (i-USEr)*, pages 18–23. IEEE, IEEE, 2011. doi: 10.1109/iuser.2011.6150529.

Oliver Niebuhr, Alexander Brem, Eszter Novák-Tót, and Jana Voße. Charisma in business speeches–a contrastive acoustic-prosodic analysis of Steve Jobs and Mark Zuckerberg. In *Proceedings of the 8th International Conference on Speech Prosody*, Boston, MA, USA, 2016.

Jillian J. M. O'Connor and Pat Barclay. The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, 2017.

Jun-Ho Oh, David Hanson, Won-Sup Kim, Young Han, Jung-Yup Kim, and Ill-Woo Park. Design of Android type Humanoid Robot Albert HUBO. In *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 1428–1433. IEEE, 2006.

John J. Ohala. Cross-language use of pitch: An ethological view. *Phonetica*, 40: 1–18, 1983.

Ian S. Penton-Voak, Nicholas Pound, Anthony C. Little, and David I. Perrett. Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition*, 24(5):607–640, 2006. doi: 10.1521/soco.2006.24.5.607.

Aaron Powers, Adam D. I. Kramer, Shirlene Lim, Jean Kuo, Sau-lai Lee, and Sara Kiesler. Eliciting information from people with a gendered humanoid robot. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 158–163. IEEE, 2005.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

Paul Robinette, Wenchen Li, Robert Allen, Ayanna M. Howard, and Alan R. Wagner. Overtrust of robots in emergency evacuation scenarios. In *Proceedings of the 11th Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '16*, pages 101–108. IEEE Press, 2016.

Najmeh Sadoughi, André Pereira, Rishub Jain, Iolanda Leite, and Jill Fain Lehman. Creating prosodic synchrony for a robot co-player in a speech-controlled game for children. In *Proceedings of the 12th Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '17*, pages 91–99. ACM, 2017.

Jelle Saldien, Kristof Goris, Selma Yilmazyildiz, Verhelst Werner, and Dirk Lefeber. On the Design of the Huggable Robot Probo. *Journal of Physical Agents*, 2(2): 3–11, 2008.

Maha Salem, Micheline Ziadee, and Majd Sakr. Effects of politeness and interaction context on perception and experience of HRI. In *International Conference on Social Robotics*, pages 531–541. Springer, 2013.

Maha Salem, Gabriella Lakatos, Farshid Amirabdollahian, and Kerstin Dautenhahn. Would you trust a (faulty) robot?: Effects of error, task type and personality on human-robot cooperation and trust. In *Proceedings of the 10th Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15*, pages 141–148. ACM, 2015.

Katarzyna Samson and Patrycjusz Kostyszyn. Effects of cognitive load on trusting behavior – an experiment using the trust game. *PloS ONE*, 10(5):e0127680, 2015.

Anara Sandygulova and Gregory M. P. O'Hare. Children's Perception of Synthesized Voice: Robot's Gender, Age and Accent. In Adriana Tapus, Elisabeth André, Jean-Claude Martin, François Ferland, and Mehdi Ammi, editors, *Social Robotics*, pages 594–602. Springer International Publishing, 2015. ISBN 978-3-319-25554-5.

Matthias Scheutz, Paul W. Schermerhorn, and James Kramer. The utility of affect expression in natural language interactions in joint human-robot tasks. In *Proceedings of the First Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '06*, pages 226–233, 2006.

Joanna Schug, David Matsumoto, Yutaka Horita, Toshio Yamagishi, and Kemberlee Bonnet. Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, 31(2):87–94, 2010. doi: 10.1016/j.evolhumbehav.2009.09.006.

Lauren E. Scissors, Alastair J. Gill, Kathleen Geraghty, and Darren Gergle. In CMC we trust: The role of similarity. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 527–536. ACM, 2009.

Olympia Simantiraki, Martin Cooke, and Simon King. Impact of different speech types on listening effort. In *Proceedings of Interspeech 2018*, pages 2267–2271, 2018. doi: 10.21437/Interspeech.2018-1358. URL http://dx.doi.org/10.21437/Interspeech.2018-1358.

Valerie K. Sims, Matthew G. Chin, Heather C. Lum, Linda Upham-Ellis, Tatiana Ballion, and Nicholas C. Lagattuta. Robots' auditory cues are subject to anthropomorphism. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, volume 53, pages 1418–1421. SAGE Publications, 2009.

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

Robert Slonim and Pablo Guillen. Gender selection discrimination: Evidence from a trust game. *Journal of Economic Behavior & Organization*, 76(2):385–405, 2010.

Eszter Strupka, Oliver Niebuhr, and Kerstin Fischer. Influence of Robot Gender and Speaker Gender on Prosodic Entrainment in HRI. In *Robot and Human Interactive Communication, 2016. ROMAN 2016. IEEE International Workshop on*. IEEE, 2016.

Rie Tamagawa, Catherine I. Watson, I. Han Kuo, Bruce A. MacDonald, and Elizabeth Broadbent. The effects of synthesized voice accents on user perceptions of robots. *International Journal of Social Robotics*, 3(3):253–262, 2011.

Angeliki Theodoridou, Angela C. Rowe, Ian S. Penton-Voak, and Peter J. Rogers. Oxytocin and social perception: Oxytocin increases perceived facial trustworthiness and attractiveness. *Hormones and Behavior*, 56(1):128–132, 2009.

Ilaria Torre, Laurence White, and Jeremy Goslin. Behavioural mediation of prosodic cues to implicit judgements of trustworthiness. In *Proceedings of the eighth International Conference on Speech Prosody 2016*. ISCA, 2016. doi: 10.21437/SpeechProsody.2016-167.

Ilaria Torre, Jeremy Goslin, Laurence White, and Debora Zanatto. Trust in artificial voices: A "congruency effect" of first impressions and behavioural experience. In *Proceedings of APAScience 18: Technology, Mind, and Society (TechMindSociety '18)*, 2018.

Jürgen Trouvain, Sarah Schmidt, Marc Schröder, Michael Schmitz, and William J. Barry. Modelling personality features by changing prosody in synthetic speech. In *Proceedings of the 3rd International Conference on Speech Prosody*, 2006.

Rik van den Brule, Ron Dotsch, Gijsbert Bijlstra, Daniel H. J. Wigboldus, and Pim Haselager. Do robot performance and behavioral style affect human trust? *International Journal of Social Robotics*, 6(4):519–531, 2014.

Gerben A. Van Kleef, Carsten K. W. De Dreu, and Antony S. R. Manstead. An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology*, 42:45–96, 2010.

Gina Villar, Joanne Arciuli, and Helen Paterson. Vocal pitch production during lying: Beliefs about deception matter. *Psychiatry, Psychology and Law*, 20(1): 123–132, 2013.

Michael L. Walters, Dag Sverre Syrdal, Kerstin Dautenhahn, René Te Boekhorst, and Kheng Lee Koay. Avoiding the uncanny valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Autonomous Robots*, 24(2):159–178, 2008.

Ze Wang, Aaron D. Arndt, Surendra N. Singh, Monica Biernat, and Fan Liu. "You Lost Me at Hello": How and when accent-based biases are expressed and suppressed. *International Journal of Research in Marketing*, 30:185–196, 2013.

Janine Willis and Alexander Todorov. First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, 17(7):592–598, 2006.

Selma Yilmazyildiz, Georgios Patsis, Werner Verhelst, David Henderickx, Eric Soetens, Georgios Athanasopoulos, Hichem Sahli, Bram Vanderborght, and Dirk

Torre, I. and White, L. (2019). Trust in vocal human-robot interaction: Implications for robot voice design. In Barkat-Defradas, M., Ohala, J., Trouvain, J. and Weiss, B. (eds.). *Voice Attractiveness: biological and social insights. Aspects of sexual, social, and charismatic evaluation in speech and conversation*. Springer.

Lefeber. Voice Style Study for Human-Friendly Robots: Influence of the Physical Appearance. In *Proceedings of the 5th International Workshop on Human-Friendly Robotics*, 2012.