2017

The impact of voice on trust attributions

Torre, Ilaria

http://hdl.handle.net/10026.1/9858

http://dx.doi.org/10.24382/438
University of Plymouth

The impact of voice on trust attributions

by

Ilaria Torre

RESEARCH DEGREES WITH PLYMOUTH UNIVERSITY

A thesis submitted to Plymouth University
in partial fulfilment for the degree of

Doctor of Philosophy

CogNovo Doctoral Training Programme

Plymouth, 30 June 2017

Acknowledgements

First of all, I wish to thank Sue Denham for creating and maintaining such a unique environment as the CogNovo Doctoral Training Centre. I could not have hoped for a better place to carry out my PhD. Thanks also to the CogNovians, it was a pleasure to collaborate and socialise with you, what a great opportunity we had! There were times in which our office was literally buzzing with ideas and excitement, I will miss this a lot. It has been a privilege to be part of this group, and I am looking forward to future visits and collaborations.

I am really grateful to Jeremy and Laurence, my main supervisors, for their guidance and constant support during these years. Thanks also go to Tony, Angelo and Caroline for additional support. Thanks also to Tom for teaching me everything about ERP analysis, and to Debora for a very successful collaboration. Also many thanks to Carolyn McGettigan for reading my thesis and for the great discussions during the viva, I am looking forward to knowing more about your new project.

I would like to thank Kevin the Great for contacting me about my Master's dissertation 2.5 years ago. I am really happy to have been able to visit Mons twice (so far) to work with you and I am looking forward to collaborating again.

Thanks to the summer school organising committee. Who would have thought that 4 people could organise an incredibly successful summer school from scratch, working on it for months while doing a PhD, and without going completely crazy? I'm glad we did it together. Thanks also to the summer school students for bearing with us and at least pretending to enjoy our event. It was great to see some of you again this year, and I hope you will keep in touch and expand our network in the future. You are the future!

I also wish to thank my housemates in NRE: Andrea and Riccardo, it was great to have a small piece of Lombardy in Plymouth. Emmanuel, thanks for the knowledge exchange. Martina, you should have started your PhD and moved in earlier, you made our house light and cheerful, I will miss you! Frank, thanks for deciding to move in with me over 3 years ago. Turns out that we make a great housemate partnership!

Thanks to all the friends who came to visit me over the years: Babi, Harriet, Gala, Valerio, Valentina, Matthew, Cecco, Ste, Pol, it was great to share some of my normality with you.

Many thanks to my friends, family and cat in Como, I miss you lots! To my nerd friends, thinking about our adventures and "chocolate ice-cream" is the best cure for PhD-induced madness and depression, I just love you.

Thanks to my oldest friend Valerio. Our badminton antics played a decisive role in getting us here. I'm looking forward to more exploring of our favourite countryside, and to seeing my name on your thesis acknowledgments soon!

Finally, the biggest thanks goes of course to Frank. I would not have made it to the end without you. You are the true pillar of CogNovo, and of my life, too.

This work has been funded through a full studentship by Plymouth University and supported by the additional training receievd through the CogNovo programme. CogNovo is funded by EU Marie Skłodowska-Curie programme (FP7-PEOPLE-2013-ITN-604764) and Plymouth University.

This work is part of the CogNovo Doctoral Training Programme and has been supervised by the director of studies Dr. Jeremy Goslin and the additional supervisors Dr. Laurence White, Dr. Caroline Floccia, Dr. Tony Belpaeme, and Dr. Angelo Cangelosi.

Abstract

The impact of voice on trust attributions

Ilaria Torre

Trust and speech are both essential aspects of human interaction. On the one hand, trust is necessary for vocal communication to be meaningful. On the other hand, humans have developed a way to infer someone's trustworthiness from their voice, as well as to signal their own. Yet, research on trustworthiness attributions to speakers is scarce and contradictory, and very often uses explicit data, which do not predict actual trusting behaviour. However, measuring behaviour is very important to have an actual representation of trust. This thesis contains 5 experiments aimed at examining the influence of various voice characteristics — including accent, prosody, emotional expression and naturalness — on trusting behaviours towards virtual players and robots. The experiments have the "investment game" — a method derived from game theory, which allows to measure implicit trustworthiness attributions over time — as their main methodology. Results show that standard accents, high pitch, slow articulation rate and smiling voice generally increase trusting behaviours towards a virtual agent, and a synthetic voice generally elicits higher trustworthiness judgments towards a robot. The findings also suggest that different voice characteristics influence trusting behaviours with different temporal dynamics. Furthermore, the actual behaviour of the various speaking agents was modified to be more or less trustworthy, and results show that people's trusting behaviours develop over time accordingly. Also, people reinforce their trust towards speakers that they deem particularly trustworthy when these speakers are indeed trustworthy, but punish them when they are not. This suggests that people's trusting behaviours might also be influenced by the congruency of their first impressions with the actual experience of the speaker's trustworthiness — a "congruency effect". This has important implications in the context of Human–Machine Interaction, for example for assessing users' reactions to speaking machines which might not always function properly. Taken together, the results suggest that voice influences trusting behaviour, and that first impressions of a speaker's trustworthiness based on vocal cues might not be indicative of future trusting behaviours, and that trust should be measured dynamically.

Authors declaration

At no time during the registration for the degree of Doctor of Philosophy has the author been registered for any other University award without prior agreement of the Graduate Sub-Committee.

Work submitted for this research degree at the Plymouth University has not formed part of any other degree either at Plymouth University or at another establishment.

This study was financed with the aid of a studentship from Plymouth University.

A programme of advanced study was undertaken under the umbrella of CogNovo (FP7-PEOPLE-2013-ITN-604764), an innovative doctoral research training funded by the EU Marie Skłodowska-Curie programme. This included week-long workshops on *Research Methods*, *Experimental Methods*, *Computational Modeling*, *Public Outreach and Social Innovation*, *Entrepreneurship*, *Scientific Mythodologies*, and *The Neural Basis For Creativity*.

Relevant scientific seminars and conferences were regularly attended at which work was often presented; external institutions were visited for consultation purposes and several papers prepared for publication.

Publications and Conference Proceedings:

Torre, I. (2014). Production and perception of smiling voice. In *Proceedings of the First Postgraduate and Academic Researchers in Linguistics at York conference (PARLAY 2013)*. York, UK.

Torre, I., Goslin, J. & White, L. (2015a). Investing in accents: How does experience mediate trust attributions to different voices? In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. Glasgow, U. K..

Torre, I., Goslin, J. & White, L. (2015b, May 23). The effect of experience on trust attributions varies between two british accents. In *EPIP4: 4th International Conference on English Pronunciation: Issues & Practices* (pp. 146–148).

Torre, I. & Loesche, F. (2016, December). Overcoming impasses in conversations: A creative business. *Creativity: Theories – Research – Applications*, *3*(2).

Torre, I., White, L. & Goslin, J. (2016, May). Behavioural mediation of prosodic cues to implicit judgements of trustworthiness. In *Proceedings of the eighth International Conference on Speech Prosody 2016*. Boston, MA, USA: ISCA.

Conference Talks, Workshop Contributions, and Grants:

PUSoP2015 Plymouth University, UK. Second prize poster award. *Investing with accented speakers: Who do we trust, the honest or dishonest, "prestigious" or "provincial" accents?* June 2015.

Off the Lip 2015 Plymouth University, UK. Round table discussion. *Manipulating the Human Mind: When experiments get dirty.* September 2015.

11^{th} **eNTERFACE2015** University of Mons, Belgium. Workshop on multimodal interfaces. *Environment-aware social agents.* September 2015.

ISCA Bursary Boston, USA. Registration fee for Speech Prosody 2016 Conference. May 2016

Santander Internationalisation Mobility Support Scholarship Travel grant. May 2016

CI Conference 2016 Cognition Institute, Plymouth University, UK. Poster presentation. *Behavioural mediation of prosodic cues to implicit judgements of trustworthiness.* June 2016.

UK Robotics Week Centre for Robotics and Neural Systems (CRNS), Plymouth University, UK. Talk and demonstrations. *Investment game with NAO.* July 2016.

ColLaboratoire2016 Plymouth University, UK. Event organisation and workshop leader. *NAO that's what I'm talking about.* August 2016.

Off the Lip 2016 Plymouth University, UK. Live demonstration. *Would you trust this robot?* October 2016.

UKCreativity Edinburgh Napier University, UK. *Framing the impasse: a conversational-creative approach.* May 2017.

PUSoP2017 Plymouth University, UK. Best poster award. *Perseverance in game theory: people just don't give up on nasty game partners.* June 2017.

ExLing 2017 Heraklion, Greece. Oral presentation. *Playing the trust game with talkative robots*. In workshop *Using robots to study speech.* June 2017.

External Contacts: Ilaria.Torre@CogNovo.eu

Word count of main body of thesis: 48939

Signed: ______________________________

Date: ______________________________

Contents

List of Figures

List of Tables

1 Introduction

Trust is a fundamental aspect of human communication. In fact, the mere existence of communication is granted by a mutual trust between interlocutors. Such a natural tendency to trust speech is mediated by heuristics that give us indicators about when the speaker might not be trustworthy. For example, not being able to tell a speaker's background, prosodic indicators of aggression or dominance, or vocal identifiers of a particular social group, can all contribute to this impression. Empirical evidence indeed suggests that we make immediate trust judgements based on such vocal features. However, these findings generally rely on explicit measures, and questioning listeners about their attitudes to a voice might also affect their attitude, or cause them to express an attitude that they did not have or would not normally act on. Implicit measures, on the other hand, allow to determine attitudes to voices through actual behaviour, which should be a more reliable indicator of trust. They also allow us to examine whether voice-based prejudices survive experience of actual behaviour, which might not be expected, since these prejudices are presumably overwritten as soon as other evidence is available. All these issues are empirically examined in this dissertation, with experiments collecting implicit measures of trust towards different voices as it develops with behavioural experience. How do initial judgments change with exposure to a speaker's behaviour? Are some voices more trustworthy in certain contexts, and other voices in other contexts?

1.1 The importance of trust for speech and of speech for trust

Human communication, including vocal communication, needs trust to function. When we ask a stranger for directions, we trust that they will give us the correct information, to the best of their knowledge. The supermaxim of quality in Grice's Cooperative principle also highlights this norm, stating: "Try to make your contribution one that is true" (Grice, 1975). Trust is important even without using lexical verbal communication. For example, when we hear someone crying, we trust that something must be wrong, and we go in their aid. Evolutionarily speaking, such an environment, where everyone is expected to be trusting and trustworthy, facilitates the emergence of deceptors (e.g. Doebeli, Hauert & Killingback, 2004). However, as will be later discussed in greater detail (Section 2.1.2), deceptors are only at an advantage if there are few of them in a community. And even then, once a deceptor is unmasked, its advantage will vanish, just like the "boy who cried wolf". Thus, trust is essential for spoken communication.

The spoken channel is also the main mode of communication for people, so we must have developed a way to accurately signal our trustworthiness to others, and to detect theirs. Indeed, evidence shows that people are able to make fast and accurate trustworthiness judgments upon hearing someone's voice for the first time (McAleer, Todorov & Belin, 2014). Trustworthiness can be inferred from many vocal features, such as accent (Lev-Ari & Keysar, 2010), prosodic cues (N. Miller, Maruyama, Beaber & Valone, 1976) or emotional expressions (Schug, Matsumoto, Horita, Yamagishi & Bonnet, 2010). What is less clear is how initial impressions based on these cues will change upon repeated interaction with this voice, and with the behaviour of its owner.

1.2 Explicit vs. implicit measures

Work on trust attributions is generally based on data gathered through surveys or questionnaires, which typically involve participants rating someone's trustworthiness on a Likert-type scale. This type of data is called "explicit", because it requires the respondent to make an overt statement about trust in response to a direct question (Greenwald, 1990). However, such explicit data do not always directly correlate with behavioural data (De Houwer, 2006; Greenwald, 1990; Greenwald & Banaji, 1995; Wicker, 1969). This means that people who rate a voice as sounding "100% trustworthy" are not actually going to trust its advice 100% of the time. Talking specifically about trust, Glaeser, Laibson, Scheinkman and Soutter (2000) argue that explicit survey questions about trusting behaviours do not predict behaviours in economic games such as the investment game. This lack of correlation has been reported also for measures taken from the same subjects (De Houwer & Bruycker, 2007; Franck, De Raedt & De Houwer, 2007; Glaeser et al., 2000; Greenwald, McGhee & Schwartz, 1998; Schimmack & Diener, 2003). In a review of studies that measured attitudes and behaviours at separate times, Wicker (1969) also found that the two are not related. From this review, it appears that this non correspondence between behaviour and attitudes has been known in the social psychology literature for a long time, yet researchers have been mostly using explicit data, rather than behavioural data. How to define and examine the implicitness of implicit measures Implicit measures, instead, estimate attitudes towards objects indirectly, through processes that are "uncontrolled, unintentional, goal-independent, purely-stimulus-driven, autonomous, unconscious, efficient, or fast" (De Houwer & Moors, 2007, p. 192). Implicit measures can evoke attitudes through actual behaviour, where attention is not directly drawn to the particular object of interest to the experimenter (see Schwarz, 1999). Examples of implicit measures are interpersonal distance, which can be used to study openness and positive attitudes towards an attitude object (e.g. Haring, Matsumoto & Watanabe, 2013), emotional expressions, which can reveal how a participant is feeling towards something else (Oswald, Mitchell, Blanton, Jaccard & Tetlock, 2013), or economic games, which provide implicit measures of trusting behaviours and perceived trustworthiness (Berg, Dickhaut & McCabe, 1995). These will be discussed in detail in Chapter 3. Arguably, the most known implicit measure is the IAT, ideated by Greenwald et al. (1998). This test uses reaction times to measure automatic evaluations of a class of attitude objects, by matching a target concept (for example, "young") with a description (for example, "pleasant"). The IAT is believed to not be influenced by factors such as social desirability bias – where people might tweak their responses in order to appear in a certain way (Banse, Seise & Zerbes, 2001). However, even the IAT is not exempt from flaws. For example, it has been shown that it might also not always correlate with behaviour (for a review, see Oswald et al., 2013), and that participants often realise that they are being primed towards a certain attitude object (De Houwer, 2006).

1.3 First impressions

First impressions, such as those that questionnaires or the IAT typically measure, are very important, because they contribute to determining future attributions, but they do not reveal anything about any subsequent attitude development. As Solomon Asch put it, “We look at a person and immediately a certain impression of his character forms itself in us. A glance, a few spoken words are sufficient to tell us a story about a highly complex matter. We know that such impressions form with remarkable rapidity and with great ease. Subsequent observations may enrich or upset our view, but we can no more prevent its rapid growth than we can avoid perceiving a given visual object or hearing a melody” (Asch, 1946, p. 258). Thus, even a definition of first impressions contains some information about “subsequent observations”. Attitude development ought, for maximum utility, to reflect actual experience rather than remain based on initial prejudices, which might later prove unfounded. Yet, most studies on personality traits limit their investigation to first impressions, without taking into account how the attributions might evolve over time.

A study looking at the trustworthiness of Embodied Conversational Agents (ECA) partially addressed the question of trustworthiness attributions over time (Elkins & Derrick, 2013). They hypothesised that trust is temporally variant, so they collected trustworthiness ratings at different points during the interaction with the agent. However, the agent randomly changed its appearance (female or male, smiling or neutral) after each rating, so it is possible that participants simply rated what they perceived to be a new agent from scratch, rather than the same one over time. Elkins and Derrick (2013)’s approach still used explicit measures, but it is a starting point to address the temporal aspects of trust formation.

Explicit measures such as questionnaires are often administered once, and therefore only provide data about first impressions, not the dynamics of the perceiver-attitude object relation over time. In real life however, personality attributions such as trust evolve continuously, based on the behaviour of the interacting parties. In fact, one of the characteristics that seem to be common to definitions of trusting behaviour is that it increases incrementally via repeated positive interactions (Weber, Malhotra & Murnighan, 2004; Jonker & Treur, 1999; Castelfranchi & Falcone, 1998; G. R. Jones & George, 1998). For example, according to Gambetta (1988), trust "is a particular level of the subjective probability with which an agent assesses that another agent or group of agents will perform a particular action". Thus, at each interaction, a perceiver will give a value to the trustworthiness of the interaction partner at that particular point in time. Then, the perceiver's subjective probability of the partner's trustworthiness will change according to this value at the next interaction. Time, then, is likely to play a role in the shaping of trustworthiness attributions, but we still have to understand whether first impressions will gradually fade out as the object becomes more familiar or, on the contrary, will reinforce themselves due to a perceived negative behaviour (Asch, 1946).

1.4 Context dynamics

Very little research has been dedicated to experimentally manipulating the behavioural characteristics of an agent in order to examine their effect on the trust development process, despite the general agreement that attitudes are formed in time and can change based on a person's behaviour (Weber et al., 2004; King-Casas et al., 2005; Asch, 1946; Lewicki, Tomlinson & Gillespie, 2006). In particular, the effect of experience of the other party's behaviour remains unclear. Recent work in Human-Machine Interaction aimed at automatically predicting personality traits from multimodal cues partially addressed this question. Celiktutan and Gunes (2016) and Joshi, Gunes and Goecke (2014) asked participants to judge pre-recorded conversations between a human and an avatar which exhibited different emotional and personality characteristics, like sadness or happiness. They found that this situational context affected people's perception and ratings, although the authors did not mention how the judgments changed over time. Batrinca, Lepri, Mana and Pianesi (2012) also specifically manipulated behaviour: participants played a map task with a virtual agent which was programmed to have different levels of cooperativeness, in order to elicit different reactions from the participants. Surprisingly, they found that the different cooperativeness levels did not elicit different personality attributions, nor did they help to better predict them. However, the way the cooperativeness levels were chosen in this experiment is not clear, and their conclusion remains dubious.

Other temporal dynamics are likely to affect personality judgments as well. While it has been shown that very little time is necessary to form first impressions (Zebrowitz & Collins, 1997; McAleer et al., 2014; Willis & Todorov, 2006; Bar, Neta & Linz, 2006), it is also true that exposure time increases accuracy and confidence in personality judgements (Blackman & Funder, 1998; Ambady, Hallahan & Conner, 1999; Carney, Colvin & Hall, 2007; Willis & Todorov, 2006). Furthermore, it is also well established that interactants will adapt to each other's behaviour, gesture, and speech over time (Burgoon, Stern & Dillman, 2007; Giles & Powesland, 1997). With this continuous adaptation, behaviours dynamically change in response to affect, speaking partner, and environment (Elkins & Derrick, 2013). Evidence of attitude change also comes from Contact Theory: intergroup interactions, real or imagined, have been found to reduce initial prejudice (Pettigrew & Tropp, 2006; Crisp & Turner, 2009; Wright, Aron, McLaughlin-Volpe & Ropp, 1997; Blair, Ma & Lenton, 2001), both in adults and in children (Cameron, Rutland, Brown & Douch, 2006). Similar trends have been found for accent prejudice (Adank, Stewart, Connell & Wood, 2013). Length of relationship has also been found to affect trust and trusting behaviour (Engle-Warnick & Slonim, 2006), as well as strength of relationship (Burt & Knez, 1995). Thus, while it is clear that personality attributions change with experience, more evidence is needed from experimentally manipulating the behaviour of the attitude object.

First impressions, particularly of trustworthiness, might also be remnants of a "fight or flight" instinct, as pointed out by Bzdok et al. (2011). This would suggest that initial trustworthiness impressions should be quite accurate and remain relatively unchanged. The fact that trustworthiness judgments are made very quickly (Todorov, Pakrashi & Oosterhof, 2009; McAleer et al., 2014; Willis & Todorov, 2006) suggests that, at least in evolutionary terms, there was a need for such a rapid assessment of a newly encountered individual. In fact, people seem to be very good at recognising untrustworthy individuals, since this ability was necessary for survival (Carney et al., 2007). However, societal change in the history of humankind means that we most likely do not need to activate such a "fight or flight" response every time we hear a new speaker (Riedl & Javor, 2012). On the contrary, we have learned to observe other people's behaviour and build our impressions based on that: cooperation and trust are concepts built in time. Supporting this argument, several scholars have argued that a context-based measure of accuracy of personality perception is needed (e.g. McAleer et al., 2014; Salem, Ziadee & Sakr, 2013; Walters, Syrdal, Dautenhahn, Te Boekhorst & Koay, 2008). Whether first impressions are actual vestigial remnants of our ancestral heritage, or cultural biases, or a combination of the two, we do not know.

1.5 Contribution

According to Lewicki et al. (2006), two approaches have emerged in research on trust, one based on behavioural economics, and one on personality psychology. While these two have provided experimental methods and results on trust in many different settings independently of each other, only recently have the two started to join forces. Scholars in behavioural economics theorise about trust as a rational agent's behaviour, which calculates gains and losses of each potential action before making a decision. Such theories would often lack discussion about individual differences or normative social conventions, as will be discussed later (see Chapter 9.5). Conversely, psychology scholars studying trust would often rely on data from personality questionnaires (Riedl & Javor, 2012). These results have been shown to not predict actual behaviour, though (e.g. Glaeser et al., 2000), and as such they might not be appropriate if one is interested in behavioural measures.

Recently, a trend has emerged in the cognitive sciences to gamify psychological experiments, in an attempt at engaging participants more, so that they will be more intrinsically motivated to complete the task (Lumsden, Edwards, Lawrence, Coyle & Munafò, 2016; Sailer, Hense, Mayr & Mandl, 2017). Economic games such as the ones often used in behavioural economics, like variations of the Prisoner's Dilemma (see Section 3.1), have then been used to study actual trusting behaviours of participants in a standard, controlled laboratory environment (Cochard, Van & Willinger, 2004). Evidence has since shown that this type of data, when paired with other types of data (such as explicit evaluations or biological correlates) does indeed provide a bigger picture of the cognitive processes behind trusting behaviours (Glaeser et al., 2000; Riedl & Javor, 2012). The contribution of the gamified data is that it represents actual participants' decisions on a dilemma, rather than speculations about a decision-making process. Rather than asking participants to give a numerical evaluation on a question such as: "How trustworthy do you think people wearing hoods are?", their trust can be directly measured from behavioural choices, for example in economic games played with a hooded person. Similar attempts of measuring behaviour and increasing immersion is being used in moral decision-making, for example Francis et al. (2016). The distinction being made is between an individual's conscious attitudes, or propensity to trust, and actual decisions, which the individual might be only partially aware of (De Houwer, 2006; Uleman, Saribay & Gonzalez, 2008).

In this thesis, I add a third element to the economical-psychological study of trust, that of speech. Human interactions are generally based on verbal communication, which requires trust to function. Thus, we should be able to quickly asses another person's trustworthiness from the sound of their voice. As reviewed in Section 2.1.5, there is a substantial body of work on attitudes and explicit evaluations of different voices. However, often the emphasis of such studies has been on certain indexical information that someone's voice conveys (for example gender or place of origin), but rarely has the effect of vocal characteristics on trust been studied *per se*. The information conveyed in the voice, taken all together, might be a good measure of *person* perception, where a person's voice interacts with all sorts of other features, such as physical appearance or attire. However, not all interactions which require trust happen face-to-face, and this phenomenon can only increase with the advent of powerful and inexpensive voice-based technologies (see Riedl & Javor, 2012). Also, speaking machines are becoming a reality for an increasing number of people, and scholars have been also studying *agent* perception (e.g. Nass, Moon, Fogg, Reeves & Dryer, 1995). As technological advances are not likely to stop anytime soon, studying the effect that different agent characteristics, including voice, have on human decision-making, is paramount to ensure a smooth transition to a world where humans and machines cooperate in society. With this thesis, I hope to contribute to this transition.

1.6 Problem definition

To give a practical example of the issues addressed in this thesis, consider the following simplistic, hypothetical situation: Carmen is driving to her friend's graduation party; she has only a vague idea of how to get to the party, so she set the location address in her navigator system and is now following its directions. At an intersection, the navigator tells Carmen to go right, but she spots a street sign pointing to her destination on the left. What to do? It appears that she has two choices: either follow the navigator, trusting that it will have more information about current traffic conditions and driving times, or she can follow the sign, trusting that it will lead to her destination eventually. Now let's assume that the navigator system has a voice, for example a female, high-pitched voice with a Queen's English accent. This voice sounds competent and reliable to Carmen, so she quickly chooses to follow its advice, and turns right. It turns out that the street on the right is a dead-end, and Carmen has to go back, slightly annoyed at the navigator about the lost time. When the same situation presents itself again, with road signs and navigator pointing to different directions, Carmen decidedly follows the road sign. Let's now go back to the first intersection and assume, this time, that the navigator has a female, low-pitched Liverpool-accented voice instead. Carmen chose it to be her navigator voice because it makes her smile. When the intersection with contradicting directions appears, Carmen immediately follows the sign, almost without thinking. The different navigator systems' voices elicit different reactions from Carmen.

Assuming that people do indeed have attitudes towards machine agents such as navigation systems, or that they even treat them as agents – empirical evidence shows that they do (Large, Clark, Quandt, Burnett & Skrypchuk, 2017) – we might also venture to say that the same mistake, made by two navigators with different voices, might also elicit different reactions from Carmen. While a mistake from the first voice might have evoked particularly strong negative reactions, a mistake from the second, less competent-sounding voice might have evoked a milder reaction.

Thus, as has been pointed out a few times in this introduction, it seems that there are two outstanding questions in the study of voice-based trust attributions:

- Are actual trusting behaviours, rather than explicit trustworthiness judgments, influenced by voice cues?
- How does initial trust in a voice change with experience of a speaker's behaviour?

A methodology derived from game theory, called the "investment game", can be used to study trusting behaviours over time. This method is described in detail in Section 3.3. This will be used to address these questions in the following chapters.

1.7 Summary of chapters

In Chapter 2, literature on trust is reviewed with particular emphasis on the evolutionary importance of being able to accurately detect and signal trustworthiness. Studies on vocal indicators of trustworthiness judgments, particularly accent and prosodic cues, are also discussed, and inconsistencies in the literature are highlighted. In Chapter 3, the methodology used in the current experiments is described, with reference to its origin, game theory. Chapter 4 is the first experiment using the investment game as a method to gather implicit trust attributions to different accented speakers. Chapter 5 expands the preliminary results from the previous Chapter to an experiment where more accents and more speakers of the same accent are used, in order to examine individual prosodic cues that might implicitly influence trust. In Chapter 6, pitch and articulation rate of male and female speakers are artificially manipulated to directly examine the effect of prosodic cues on trust. Chapter 7 presents an experiment using smiling voice as a cue to trust. Chapter 8 presents an experiment using the investment game in Human-Robot Interaction, where a robot has either a synthetic or a natural voice. Finally, general conclusions on the vocal indicators of trust are drawn in Chapter 9, together with theoretical implications and potential practical applications.

2 Literature review

This chapter starts with a general introduction to the concept of trust, then describes where trust lies in the social psychology literature and the attribution of trustworthiness to voices. In the second part of this chapter, theories and experimental evidence for accent and voice attitudes are discussed. Finally, the study of trust in Human-Machine Interaction is introduced.

2.1 Trust

2.1.1 Definitions

Trust makes someone, the trustor, knowingly accept vulnerability to a trustee. More formally, it is defined as a "psychological state comprising the intention to accept vulnerability based upon positive expectations of the intentions or the behavior of another" (Rousseau, Sitkin, Burt & Camerer, 1998). Another definition by Mayer, Davis and Schoorman (1995, p. 712) emphasises the reason behind the trustor's decision and the independence of the involved people: "[trust is] the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party". From these definitions, it emerges that trust is interpersonal: there must be someone/something who trusts (a so called "trustor") and someone/something who is trusted (a so called "trustee"). The trustor has no control over whether their trust has been well-placed or not, so they are exposed to some uncertainty. Nevertheless, the trustor supposedly has some mental model of how trustworthy the trustee is (Gambetta, 1988), and this will influence his/her decision to trust or not. This is what Mayer et al. (1995) call the "willingness" to take a risk.

Other definitions of trust mention the importance of cooperation. For example, Burt and Knez (1995, p. 257) say that "Trust is anticipated cooperation", and Gambetta (1988, p. 217) says that someone is trustworthy when "the probability that he will perform an action that is beneficial or at least not detrimental to us is high enough for us to consider engaging in some form of cooperation with him". However, trust can happen without cooperation, for example in the case of children's unconditional trust. Also, cooperation does not always need trust, for example when cooperating does not put any of the involved parties at risk (Mayer et al., 1995). In these situations, trust is more marginal to the interaction. Still, in the cases in which cooperating is a risky decision for someone, cooperation can be used as an accurate measure of trust, since they are deeply intertwined. Thus, just like for risk, trust includes a willingness to cooperate, even though cooperation does not need to happen.

From these definitions, we can imply that trust is a behaviour (Boone & Buck, 2003; Fehr, 2009a; Mayer et al., 1995): the trustor's decision to trust makes him/her vulnerable to the trustee's behaviour. Thus, to make a clear connection between some key terms in the trust literature, the trustor can be trusting, and the trustee can be trustworthy. If the trustee is untrustworthy, but still tries to elicit the trustor's trust, he/she is deceptive. Deception can be achieved by telling lies, which is the opposite of telling the truth. If the trustor trusts the trustee, it means that he/she believes the trustee to be trustworthy. The trustworthiness of the trustee is both a real trait that the trustee possesses, and an attribution that the trustor makes, which can be well-placed or not. In this thesis, the term trustworthiness denotes the latter, unless otherwise specified.

2.1.2 Evolutionary theories of trust

Trust is a fundamental aspect of social interactions, to the extent that, without trust, human societies would not exist (Baier, 1986; Yamamoto, 1990; Luhmann, 1979; Lagenspetz, 1992), and indeed scholars have consistently found that low income countries generally have low trust (Zak & Fakhar, 2006; Zak & Knack, 2001; Beugelsdijk, De Groot & Van Schaik, 2004; Knack & Keefer, 1997; La Porta, Lopez-De-Silane, Shleifer & Vishny, 1997). In fact, we make trusting decisions every day, and many times per day (Luhmann, 1979). At a macro-societal level, we trust our country's institutions to provide for our needs, for example to stock our markets with food, to provide accessible education, and to defend us against discrimination. At a micro-societal level, we trust drivers to adhere to traffic laws, restaurants to not serve us spoiled food, and other parents to vaccinate their children. We also trust at individual levels, for example when we trust a friend to look after our pets, our partner to not cheat on us, and our children to not crash the car when they go out. These different levels of trust are all related to each other, and all contribute to a functioning society (Rothstein & Uslaner, 2005).

But is trust a uniquely human behaviour? Some evidence suggests that animals display some forms of trust as well (Bateson, 2000). For example, male baboons might help another male to win over a female, trusting that the favour will be returned one day (Packer, 1977). Also, bats will share their food with non-kin individuals who have not fed enough, knowing that a small sacrifice today might result in a gain tomorrow, when they are the malnourished ones (G. S. Wilkinson, Carter, Bohn & Adams, 2016). Such kind of reciprocal altruism has been described as a form of trust in the animal world, and as an evolutionary strength (Harcourt, 1991).

Thus, in both the animal and the human world, the society as a whole is better off if there is trust between individuals. However, individuals who do not cooperate, while everyone else does, are even better off, because they exploit everyone's labour without paying its cost (Axelrod & Hamilton, 1981; Trivers, 1971). Evolutionarily speaking, cooperative contexts have been demonstrated to favour the emergence of these defectors, or "free riders" (Doebeli et al., 2004). Free-riding is only beneficial to the individual if there are few individuals who do it, though. For example, classmates who do nothing on the group project, knowing that the other people in the group will do the work for them, can benefit from the final grade without putting any work into it. If everyone does that, however, the teacher will fail all the students in the group, and everyone will be worse off. Thus, there must be a balance between cooperative and uncooperative individuals (Doebeli et al., 2004), and being able to recognise uncooperative individuals is an advantage for everyone. In fact, everyone wants to interact with cooperative individuals, either to cooperate with them, or to exploit them (Boone & Buck, 2003).

Recognising potential deceivers is thus a strategy that allows the individuals, and by extension the whole society, to survive. Experimental evidence confirms that people are able to distinguish trustworthy from untrustworthy individuals very rapidly. For example, Willis and Todorov (2006) showed that very little exposure time (100 ms) is enough for people to form first impressions, including of trustworthiness, competence, likeability, attractiveness and aggressiveness. They also showed that people's impressions become more confident if the exposure time increases. Of all these traits, trustworthiness was the impression that had the highest correlation between judgements made after 100 ms and judgements made without time constraints. The authors' conclusion is that this might be because, in evolutionary terms, immediate trustworthiness detection is essential for survival. On the other hand, Carney et al. (2007) found that exposure time increased accuracy especially in the detection of positive traits, including extraversion and agreeableness — which is a major trait that includes trustworthiness, as discussed below. Instead, exposure time did not improve accuracy for negative affect, neuroticism, openness, and intelligence. They suggested that the former traits might be linked to social interactions, which might take more time to be unravelled, whereas the latter might be linked to threat and competence, for which a quick decision might be needed. In general, they found that accuracy was better for negative traits, perhaps because, in evolutionary terms, being accurate in these kinds of situation is necessary for survival. Thus, it is possible that people might need to be more quick and accurate about a person's untrustworthiness, rather than trustworthiness.

In support to the argument that people might be more accurate in recognising untrustworthy individuals, several studies found that people remember untrustworthy individuals better than trustworthy ones. For example, Yamagishi, Tanida, Mashima, Shimoma and Kanazawa

(2003) found that participants remembered the faces of players who defected in a prisoner's dilemma game better than the faces of cooperative players. Rule, Slepian and Ambady (2012) found that participants remembered faces that had been previously rated as untrustworthy better than trustworthy ones, even though nothing about trustworthiness was mentioned to them. Also, Suzuki and Suga (2010) found that, rather than faces, people remembered untrustworthy behaviour in an economic game better than trustworthy behaviour. This consistent evidence suggests that people spend more cognitive effort into remembering untrustworthy individuals, in order to avoid them in the future. Following a parsimonious principle, this suggests that the default in human interaction is to be cooperative, and that it's uncooperative individuals who "stand out" and need to be marked as different (cf G. R. Jones & George, 1998). In fact, in artificial societies where deceivers outnumbered cooperators, people remembered the cooperators better (Barclay, 2008). So, given its importance, humans must have developed a way to signal, and decode, trustworthiness.

2.1.3 Biological aspects of trust

Trust also has biological and neurological components. In fact, it has been suggested that trusting behaviours are to some extent genetically predetermined (Riedl & Javor, 2012), for example through the oxytocin receptor gene (Reuter et al., 2009). Oxytocin is a hormone that has been shown to influence prosocial behaviour in humans, like maternal attachment and pair bonding (Donaldson & Young, 2008). Oxytocin levels were found to be higher when participants in a trust game were trusted by their partners, and when they trusted their partners (Zak, Kurzban & Matzner, 2005). Nasally administered oxytocin has also been found to increase trusting behaviours in a trust game (Kosfeld, 2007). Serotonin also seems to be linked to trust, as deficiencies in this neurotransmitter resulted in reduced cooperation in ultimatum games (Crockett, Clark, Tabibnia, Lieberman & Robbins, 2008) and iterated trust games (Wood, Rilling, Sanfey, Bhagwagar & Rogers, 2006).

Studies on patients with brain lesions found that participants with damage in the amygdala tended to trust untrustworthy partners in a repeated trust game (Koscik & Tranel, 2011), and to explicitly rate faces as more trustworthy than unimpaired participants (Adolphs, Tranel & Damasio, 1998), suggesting that the amygdala might play an important role in trustworthiness detection. A meta-analysis of 16 fMRI studies by Bzdok et al. (2011) also highlighted the special function that the amygdala seems to play in trust. The authors further speculate that the amygdala might be the determinant for "fight-or-flight" decisions as well as social behaviours with long-term ends, such as trustworthiness. A review by Riedl and Javor (2012) determined that literature on the neural correlates of trust is quite consistent in its findings across different experimental paradigms. In particular, they distinguish five classes of mental processes that are crucial in trust situations, and map them to different brain regions: reward processing (striatum and thalamus); risk (amygdala, insular cortex, hippocampus and parahippocampus gyrus); memory (necessary for weighing trust at each interaction based on previous experience — amygdala, hippocampus and parahippocampus gyrus); processing of cognitive conflict (cingulate cortex); mentalising (frontal cortex).

Mentalising, also known as Theory of Mind (ToM), plays an important role in trust situations, since it is one of the ways that help trustors decide whether to trust or not: by thinking about the trustee's mental processes, trustors can derive his/her intentions to be trustworthy or not (Fehr, 2009b; Riedl & Javor, 2012). Upon meeting a new person, many people will assume that he/she can be trusted, simply because it is easier and more statistically accurate than to distrust them from the beginning (G. R. Jones & George, 1998; Barclay, 2008), before their relative trustworthiness is quickly assessed (Todorov et al., 2009). However, individual differences play an important role in someone's propensity to trust (Glaeser et al., 2000). For example, an individual's moods or attitudes will affect their trust towards certain other individuals (G. R. Jones & George, 1998). Also, certain people start from the opposite assumption that no-one can be trusted, and for this reason they themselves are very often untrustworthy. These people are sometimes called Machiavellians (Repacholi, Slaughter, Pritchard & Gibbs, 2003; D. N. Jones & Paulhus, 2009), from the Italian Renaissance writer Machiavelli, who argued that deceptive behaviour is an effective political tool. Apart from generally faring better than non-Machiavellian people in economic games based on trust (Czibor & Bereczkei, 2012), it seems that Machiavellians activate different brain structures during the trusting decision process. In particular, Machiavellian individuals show less activity in regions linked with social empathy, such as the medial and ventromedial prefrontal cortices, the inferior parietal cortex, and the superior temporal sulcus. Instead, they activate regions linked with inference making and reward-related decision making, such as the inferior and middle frontal gyrus, the anterior insula, the thalamus and the anterior cingulate cortex (Bereczkei, Deak, Papp, Perlaki & Orsi, 2013).

The existence of genetic and neural factors linked to trust, and the presence of some forms of trust in animal societies suggest that trust has been an essential aspect of human interactions for a very long time. Thus, humans must have evolved some way of quickly decoding the trustworthiness of a newly encountered individual, as well as a way of signalling their own trustworthiness. In fact, some theories even hypothesise that it was the need for a cooperative society that started the process of language in our simian ancestors (Knight, 1998), and indeed language is an extremely cooperative process by itself (Tomasello, 2008). Regardless of the controversial theories on the origin of language, given that spoken language is our main mean of communication, we must be able to signal and detect trustworthiness from voices. This thesis is an investigation on what characteristics of the human voice might elicit trust or distrust in different contexts.

2.1.4 Trust and personality traits

In personality psychology, trusting and trustworthiness have been studied as personality traits. In the classical 5-Factor Model, trust is located in the agreeableness domain, which, together with openness to experience, conscientiousness, extraversion and neuroticism constitutes the “Big 5” (McCrae, 2009; Pervin, 2001). The agreeableness domain is associated with interpersonal traits such as warmth, friendliness, altruism and compliance to the needs of others (Digman, 1990; Graziano & Tobin, 2002; Graziano & Eisenberg, 1997).

Agreeableness is also related to Theory of Mind, in particular to social perceptual ToM (Nettle & Liddle, 2008). This refers to the process of inferring someone’s mental state from primary information such as eye gaze and facial and vocal expression, as opposed to social-cognitive

ToM, which is the process of using the reasoning about someone's mental state to predict their behaviour (Tager-Flusberg & Sullivan, 2000; Amodio & Frith, 2006). People who are better mentalisers often behave as they would like other people to behave with them (Baron-Cohen, 2009; Shamay-Tsoory, 2011). Attributing some personality traits to others, or having attitudes towards others, is then also informed by perceptual ToM (for example through physical appearance, attire, indexical information from the voice, etc.). Thus, we can argue that trusting behaviour will be modulated by a social perceptual/attitudinal stage first, when first impressions are formed, and by a social cognitive stage second, when first impressions drive the speculation about the trustee's future behaviour.

Assessing someone's trustworthiness might be done by broadly categorising that someone is a "good guy" or a "bad guy" (Todorov & Duchaine, 2008), which informs the perceiver about whether the best strategy is approach or avoidance (Chen & Bargh, 1999; Cosmides & Tooby, 2000; McAleer et al., 2014). Thus, people showing some positive traits are often attributed other positive traits, and people showing negative traits are attributed more negative traits, in a sort of "halo" effect (Nisbett & Wilson, 1977; Wetzel, Wilson & Kort, 1981). Notably, physical attractiveness is particularly strongly associated with other positive traits, so much that Dion, Berscheid and Walster (1972) coined a "what is beautiful is good" stereotype (e.g. Jackson, Hunter & Hodge, 1995; Mazzella & Feingold, 1994; Patzer, 1983; Langlois et al., 2000; Zebrowitz & Montepare, 2008). Attractiveness and trustworthiness seem to be deeply connected (Todorov, Said, Engell & Oosterhof, 2008), even at the neural level (Bzdok et al., 2011). Bzdok et al. (2011) concluded that specific brain regions, such as the amygdala, might act as a filter for sensory information with big evolutionary importance, such as long-term social information (e.g.: "Is this person attractive? I might date them in the future" and "is this person trustworthy? I might collaborate with them in the future").

2.1.5 Trust in the voice

A "halo" effect might also apply to attractiveness perceived from the voice. In fact, Zuckerman and Driver (1989) found that vocal attractiveness elicited higher judgments of dominance, achievement and likeability, and proposed a “what sounds beautiful is good” stereotype. This is corroborated by other studies. For example, in a study on impression formation of politicians, Surawski and Ossoff (2006) found that vocal attractiveness elicited other positive judgments, including of trustworthiness. Attractive voices were also more persuasive, in both explicit and implicit measurements (Chaiken, 1979). Vocally attractive individuals were rated higher in job interviews and were reported as having better job performances (DeGroot & Kluemper, 2007), and were more persuasive leaders (DeGroot, Aime, Johnson & Kluemper, 2011). Additionally, speakers were rated as more persuasive, competent and sociable if they had an attractive voice (Burgoon, Birk & Pfau, 1990).

But what does it mean for a voice to be attractive? In most of the research on vocal attractiveness, people were asked to rate a range of voices on an attractiveness scale, and then the most and least attractive voices were used for a personality rating test. Very few studies actually tried to map attractiveness onto an acoustic space. According to a preliminary study by Zuckerman and Miyake (1993), an attractive voice is “intermediate in its loudness and more resonant” [p. 128]. DeGroot and Motowidlo (1999) argued that attractive voices for managers had faster speech rate, fewer pauses, lower variability in loudness, lower pitch and higher pitch variability. The attractive voices in Burgoon et al. (1990) were fluent and exhibited greater pitch variability. Also, Bruckert et al. (2010) found that morphing several voices into one makes the voice thus obtained more attractive, perhaps because it eliminates small audible differences that deviate from the mean. The only in-depth study on the acoustics of vocal attractiveness is Babel, McGuire and King (2014), who found that several features affected the attractiveness of male and female voices. In particular, male voices were rated as more attractive when they had lower first formant frequencies for the /i/ and /u/ vowels, and when they exhibited shorter duration. By contrast, the most attractive female voices were breathy and showed u-fronting, which is typical of young Californian women, suggesting that group-membership might play a role in attractiveness perception as well.

Babel et al. (2014)'s finding on u-fronting raises the issue of the vocal attractiveness of accents. However, most of the research on accent evaluation includes a "social attractiveness" trait (e.g. Bishop, Coupland & Garrett, 2005), which is arguably different from a "physical attractiveness" one. While the latter has been described as "sounding good" (e.g. Zuckerman & Driver, 1989), the former is more linked to prestige (e.g. Bishop et al., 2005). Stemming from the "halo" effect described above, it seems likely that these two attractiveness traits will be correlated. Thus, vocal attractiveness and social attractiveness (and trustworthiness, as discussed above) might be related. A more detailed discussion on the accent literature follows, in Section 2.2.1.

Given that the main purpose of attractiveness is to find mates (Grammer, Fink, Møller & Thornhill, 2003; Rhodes, 2006; Hughes, Dispenza & Gallup, 2004), vocal attractiveness is strongly related to sexual orientation, and appears to be mostly gender-specific. In fact, females consistently rate low-pitched male voices as more attractive, while males rate high-pitched female voices as more attractive (Re, O'Connor, Bennett & Feinberg, 2012; Saxton, Caryl & Roberts, 2006; Hodges-Simeon, Gaulin & Puts, 2010; O'Connor & Barclay, 2017; Riding, Lonsdale & Brown, 2006; Borkowska & Pawlowski, 2011; Feinberg, DeBruine, Jones & Perrett, 2008; B. C. Jones, Feinberg, DeBruine, Little & Vukovic, 2010). From the same evolutionary argument, it can be evinced that very raucous or creaky voices might not be perceived as very attractive, since they can be a sign of poor physical health (Blood, Mahan & Hyman, 1979; Laver, 1968). On the other hand, breathy voices in women are linked to desirability (Henton & Bladon, 1985), and are often rated as attractive (Xu, Lee, Wu, Liu & Birkholz, 2013; Liu & Xu, 2011; Babel et al., 2014). More on the vocal attractiveness/trustworthiness will be discussed in Section 2.2.2.

Thus, the literature on the effects of voice on personality attributions seems to be limited to work contexts — e.g. the managers in DeGroot and Motowidlo (1999), the job interviews in DeGroot and Kluemper (2007) and the leaders in DeGroot et al. (2011) — and mating context — e.g. all the findings on sex-mediated preferences for high and low pitch mentioned above. In these contexts, it suggests that certain vocal features, in particular those related to attractive voices, are also correlated with trustworthiness. However, these studies were mostly focused on the effect of certain voice types, and only looked at the voices post-hoc. What is lacking is a systematic analysis of different voice types, differing in a few carefully balanced characteristics, and their effect on personality attributions. This was partially achieved by some of the studies on partner selection on the basis of voice, which manipulated the pitch of male and female voices and collected various ratings of trustworthiness and attractiveness, among other traits (e.g. O'Connor & Barclay, 2017; Re et al., 2012). However, these studies only focus on partner selection procedure, and thus they typically only manipulate one vocal characteristic (pitch). But the human voice is much more than pitch. Features such as accent, pitch range, and voice quality, just to name a few, contribute to an individual's voice at a particular moment in time. Different combinations of these features carry a range of subtle information and are likely to result in distinctive personality perceptions. This is what I contribute to unravelling in the experiments presented in this thesis.

2.2 Voices

2.2.1 Accents

Accents can be defined as any systematic differences in pronouncing the sounds of a language, that people belonging to a certain group (geographic or social) have in common (Lippi-Green, 1997). Accents are distinct from dialects, which are systematic differences in terms of morphology, syntax, lexicon and pronunciation of a language (Trudgill, 2000). They are also different from idiolects, which are an individual's specific use of language — as opposed to a group's (Bloch, 1948). From this point of view, every speaker has an accent, although some accents might be less marked and more desirable than others, for a variety of reasons (Lippi-Green, 1997; Niedzielski & Preston, 2000).

A distinction also needs to be made between native accents — for example, Cockney, Received Pronunciation (RP) and Scottish are accents of English — and non-native accents — such as those of mother-tongue Indian or Japanese people speaking in English as their second language. Native accents of a certain language can have allophonic contrasts that are realised differently in other accents. For example, some accents in the South West of England, Scotland and Ireland are rhotic, i.e. they pronounce a certain realisation of /r/ when it occurs post-vocalically, while others, such as RP, do not. Also, intervocalic /t/ can be pronounced as an alveolar stop (e.g. in RP), as a glottal stop (e.g. in some Northern English varieties), or as an alveolar tap (e.g. in some North American varieties). Pronunciation differences are also present in vowels, with phenomena such as u-fronting in Southern English accents or monophthongisation in Southern American English. On the other hand, non-native accents (L2) result from merging the phonological system of one's native language into the target language (Lippi-Green, 1997). Thus, Italian native speakers, who have 7 vowels in their phonological system of origin, will tend to approximate the numerous English vowels to their closest neighbour in the Italian phonological space. This results, for example, in the words "cheap" and "chip" being pronounced in the same, or in a very similar, way. Accents are part of people's identity, and they contribute to shaping our relationship with the rest of the world with whom we communicate (Fuertes, Gottdiener, Martin, Gilbert & Giles, 2012). Given that every speaker has an accent, and that accents affect the way we communicate with each other, a question arises: should speaking machines have human accents?

2.2.2 Perceptual evaluations of accents

With increased means for travelling and living away from one's home town, people are finding themselves surrounded by an increased variety of accents. As a consequence, people might be more aware of the fact that the same language can be spoken in different ways. Hearing someone's accent is unavoidable: when we hear someone speaking, we can immediately tell if he/she speaks like us or not. Unconsciously, this places the speaker either inside the group where we believe we belong, or outside of it. This in-group vs. out-group localisation then brings forward a series of attitudinal responses towards that speaker (e.g. Kinzler, Shutts, DeJesus & Spelke, 2009). As previously mentioned (Section 1.4), these initial impressions will most likely be shaped by subsequent behavioural experience with the speaker, although the first impression might not be easily abandoned.

Accents are often linked with personality stereotypes (Preston, 1999). To give some popular examples, American films and children cartoons often give their villains a British English accent (Dobrow & Gidney, 1998), and comedies will often emphasise native and non-native accents for humorous purposes (Dave, 2013). Past studies have also concluded that native accents may be perceived as more trustworthy than non-native accents (Lev-Ari & Keysar, 2010). Lev-Ari and Keysar (2010) tested whether statements uttered by non-native speakers of English with mild to heavy foreign accents would be perceived as less truthful than statements uttered by native English speakers, and found that non-native accents negatively influenced truth judgments. A preference to listening to one's native language emerges from the very beginning of someone's life (Nazzi, Bertoncini & Mehler, 1998; Moon, Cooper & Fifer, 1993), and it has been hypothesised that this preference might serve to acquire culture-specific knowledge (Marno et al., 2016). A native accent preference emerges around 4-5 years of age (Kinzler & DeJesus, 2013; Kinzler, Corriveau & Harris, 2011). Later on, this implicit preference shifts to an in-group or familiarity preference for the accent of one's social group (Abrams & Hogg, 1987; Cargile & Giles, 1997; Dragojevic & Giles, 2014). For example, Californian English was evaluated more positively by American Southern English speakers than Punjabi English (Dragojevic & Giles, 2014). Similarly, participants from Dundee evaluated Dundee speakers more positively than Glasgow speakers, and Glasgow speakers more positively than RP speakers (Abrams & Hogg, 1987). While the Dundee speakers had an overall preference for speakers of their own accent, they also preferred Glasgow to RP, perhaps because they felt that a Glasgow accent was more similar (both socially and geographically) to their own, rather than an RP accent. This seems to suggest that the concept of "in-groupness" might be concentric (cf. Clarke & Garrett, 2004, who proposed that accents can be ranked on a perceptual scale according to their distance from one's own accent). Similarly, speakers of South Welsh and Somerset varieties of English showed a tendency to "accent loyalty" for "informal" traits related to good looks and humour, while they rated the RP accent higher in "formal" traits such as competence (Giles, 1971).

A lot of research has been conducted on attitudinal evaluations of accents within the context of the British Isles (P. Garrett, Coupland & Williams, 1999; Giles, 1970; Hiraga, 2005). In particular, from such studies it has emerged that standard accents such as Standard Southern British English (SSBE) tend to be rated as more pleasant and prestigious than regional accents in general (Bishop et al., 2005; Dixon, Mahoney & Cocks, 2002; Fuertes et al., 2012; Giles, 1970; A. M. Wilkinson, 1965). For example, in a meta-analysis of 20 studies on accent evaluations, Fuertes et al. (2012) found that speakers of a standard accent are rated more positively than speakers of a non standard accent. Furthermore, standard-accented speakers generally are rated higher in terms of solidarity, a personality dimension that includes trustworthiness (e.g. Giles & Billings, 2004). Within the regional accents, countryside accents like Yorkshire or Devon are often rated as more friendly and trustworthy than city accents like London, Liverpool or Birmingham (Bishop et al., 2005; Giles, 1970; Hiraga, 2005; Kristiansen, Eiser, Harding & O'Looney, 1983; A. M. Wilkinson, 1965; Strongman & Woosley, 1967). In the biggest study so far on the topic, Bishop et al. (2005) examined the responses of more than 5000 British people to a questionnaire on the attractiveness and prestige of 34 English accents, 30 years after Giles (1970)'s seminal study. In general, they found that Received Pronunciation (RP) was still considered highly prestigious and attractive, while accents such as Liverpool and Birmingham were located at the low end of the prestigious and attractiveness scales. However, most of these studies are so called "conceptual studies" — that is, participants are given an accent label and asked to rate this label on an attribute scale. Since accents usually index geographical locations (with the exception of accents associated with social class, such as SSBE), such conceptual evaluations of accent labels are likely to be influenced by stereotypes based on general socio-economic perceptions of particular regions, rather than directly deriving from acoustic-phonetic features of the accent (Bishop et al., 2005; P. Garrett, Williams & Evans, 2005; Giles, 1970). Thus, the conclusions of previous studies based on accent labels might be more explicative of these socio-economic perceptions (Giles, 1970; Ladegaard, 1998; Omdal, 1995). However, there is evidence that people might not be very effective at localising accents upon hearing them (Bayard, Weatherall, Gallois & Pittam, 2001; Goggin, Thompson, Strube & Simental, 1991; Gooskens, 2005; Clopper & Bradlow, 2008). Thus, the question of whether people have the same attitudinal responses when hearing an accented speaker, without knowing the accent label, remains.

Context also seems to play a role in accent attitudes. For example, Wang, Arndt, Singh, Biernat and Liu (2013) found that, in a favourable context (for example, a satisfactory customer/employee call centre interaction), American interactants tended to suppress their negative prejudices of Indian-accented employees. On the other hand, in an unfavourable context (when the interaction was not satisfactory for the costumer), customers tended not to suppress their accent prejudices. Bresnahan, Ohashi, Nebashi, Liu and Shearman (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. They found that the “friend” context was judged as more attractive and dynamic than the “teaching assistant” context, in all accent conditions; and that strong and weak national identity played a role in evaluating the native and non-native accents.

There is evidence that even non-native speakers have accent attitudes. For example, in a study carried out in Denmark (Ladegaard, 1998), Danish participants rated British, American and Australian English differently, with Australian being rated as the most reliable, American as the most humorous, and British as the most prestigious. Furthermore, participants exhibited a pattern of evaluations to different British English accents similar to that found in British English participants: RP was rated higher (in terms of status and competence) than Scottish and Cockney English. Similarly, Spanish and Basque native speakers rated themselves as having more positive attitudes towards several varieties of British English (including RP) over American and Irish English (Cenoz & Lecumberri, 1999). Certain accents or idiolects seem to evoke attitudes of “prestige” because they are associated with certain social classes (Giles, 1970), and there is evidence that this type of information might be processed from specific, language-independent voice characteristics as well. For example, Brown, Strong and Rencher (1975) recorded speakers representative of various French Canadian social class levels, and listeners who did not speak French, nor did they have any knowledge of the French Canadian culture, were still able to accurately recognise the speakers' social class. This evidence contributes to suggesting that accents might be inherently coloured with phonetic-prosodic characteristics that evoke certain attitudinal responses, even to people who are not aware of the indexical information carried by accents.

Many of the seminal works on accent attitudes which did not use conceptual evaluations used the "matched-guise" technique (Lambert, Hodgson, Gardner & Fillenbaum, 1960), which consists of a bidialectal speaker reading passages in both accents, so that listeners' evaluations will not be affected by individual voice cues. However, in everyday encounters, people constantly and implicitly judge speakers by their voice, which is made of many indexical and other characteristics combined. Likely, such implicit judgments are not the sum of the judgments of each individual voice characteristic, but rather various aspects of the voice interact with each other and with listeners' individual differences, to result in what will be the listener's attitude (Kriengwatana, Terry, Chládková & Escudero, 2016). For example, a high-pitched Birmingham speaker might elicit different attitudes than a low-pitched Birmingham speaker, or a high-pitched Liverpool speaker. Thus, it becomes important to consider these and other variables when trying to determine, for example, what makes a voice sound trustworthy.

2.2.3 Personality attributions to voices

Apart from accent, other segmental and suprasegmental characteristics contribute to an individual's unique voice. While accent, as discussed above, carries information about someone's nationality or social status, other features can convey information about someone's physical appearance, age, gender, personality, emotional state, and even sexual orientation (e.g. Mack & Munson, 2012). As our primary means of communication, it seems logical that we are "tuned" to convey and decode all this information in the speech stream. Still, it is rather remarkable that we can do all this with relatively small articulators, and with a limited set of phonemes.

The fact that we are able to convey all this information in our speech suggests that we must be able to understand it as well. And in fact, evidence shows that people are able to successfully decode this information from speech, even after a very short exposure time (McAleer et al., 2014). For example, McAleer et al. (2014) administered a questionnaire on-line, in which they asked to rate the word “hello”, pronounced by 64 Scottish speakers, in terms of various personality traits, and found that listeners were consistent with their evaluations. Nonetheless, it is also believed that exposure time increases accuracy in personality judgements (Blackman & Funder, 1998; Ambady et al., 1999; Carney et al., 2007; Willis & Todorov, 2006). Thus, vocal characteristics can convey information about personality traits as well. The following section provides a review of studies dealing with traits that can be ascribed as similar to trustworthiness, such as charisma, persuasion, deception, leadership, etc., in the absence of a conclusive set of studies on trustworthiness itself.

One characteristic that has been examined in the past is speech or articulation rate, which is a measure of how fast someone speaks. While the term “speech rate” is generally used to indicate the number of speech units in a certain length of time, including pauses, “articulation rate” denotes the number of speech units articulated over a total speaking time, thus not including pauses (Jacewicz, Fox, O’Neill & Salmons, 2009; De Jong & Wempe, 2009). For example, B. L. Smith, Brown, Strong and Rencher (1975) found that speakers with a slow speech rate were rated as less competent. Similarly, Apple, Streeter and Krauss (1979) found that speakers with slow speech rate were rated as “less truthful” and “more passive”. N. Miller et al. (1976) also found that faster speech rate increased persuasion. Fast speech rate also elicited judgments of competence (Street, 1984; Street & Brady, 1982) and credibility (Buller, LePoire, Aune & Eloy, 1992), although this effect decreased for very fast rates of speech (Buller et al., 1992). Similarly, others found that a fast speaking rate is a feature of charismatic (Jiang & Pell, 2017), confident (Rosenberg & Hirschberg, 2005) and persuasive (Chaiken, 1979) speakers. On the other hand, Niebuhr, Brem, Novák-Tót and

Voße (2016) argued that fast speaking rate hinders charismatic speech, possibly because phenomena such as vowel reduction and deletion become more likely.

Pitch has also received considerable attention in the past. A perceptual correlate of the acoustic fundamental frequency (f0), the rate of vibration of the vocal chords in the larynx. Thus, this feature is at least partially determined by physiology: people who have bigger larynges will generally have a lower pitch, and vice versa. The relation between pitch and trustworthiness is rather controversial, also because of the different technologies and methodologies used in previous studies. For example, Hughes, Pastizzo and Gallup (2008) found a positive correlation between pitch and perceived honesty of male speakers. On the contrary, in the context of male political trustworthiness, participants typically voted for male candidates with lower-pitched voices in a mock election scenario (Tigue, Borak, O'Connor, Schandl & Feinberg, 2012), and it was found that low pitch predicted actual election results (Banai, Banai & Bovan, 2017). Similarly, in a questionnaire study, Apple et al. (1979) showed that speakers with a high f0 were rated as "less truthful". Elkins and Derrick (2013) also found that participants gave low trust attributions when interviewed by embodied agents that had a high-pitched voice. Regarding female political trustworthiness, female candidates with high pitch voices were more successful than those with low pitch (Klofstad, 2015). Rosenberg and Hirschberg (2009), instead, found that pitch was positively correlated with ratings of speaker's charisma. Furthermore, in an imagined version of the investment game, where participants selected which of two possible vocal partners they would hypothetically trust more, O'Connor and Barclay (2017) found that they consistently selected the higher-pitched one. Conversely, a few studies have found that participants raise their vocal pitch (f0) when lying (Anolli & Ciceri, 1997; Streeter, Krauss, Geller, Olson & Apple, 1977; Villar, Arciuli & Paterson, 2013), while Zuckerman, DeFrank, Hall, Larrance and Rosenthal (1979) found that participants perceived a lower pitch in deceptive messages, and Kirchhübel and Howard (2013) failed to find any acoustic differences — including in f0 — in the production of deceptive and truthful messages. From a slightly different perspective, Cheang and Pell (2008) and Rao (2013) found that actors generally had a lower f0 and slower speech rate when acting sarcastic voices than when they were acting sincere voices.

Similarly, sincerity in a synthetic voice was associated with greater pitch range and faster articulation rate (Trouvain, Schmidt, Schröder, Schmitz & Barry, 2006). Higher f0 was also associated with higher agreeableness (Imhof, 2010), one of the "Big Five" personality traits (McCrae, 2009; Pervin, 2001; Trouvain et al., 2006), and elicited higher cooperativeness ratings (Knowles & Little, 2016).

It should also be noted that although there is a confound between pitch and gender, with females typically having a smaller larynx and thus a higher pitch, there are no consistent differences in trustworthiness judgments to males and females (e.g. Nass & Brave, 2005; Chaudhuri & Gangadharan, 2007; Boenin & Serra, 2009; Slonim & Guillen, 2010). For example, in Elkins and Derrick (2013)'s study, there were no differences in perceived trustworthiness of female or male agents. Inconsistencies were also corroborated in a series of studies using economic games. For example, Chaudhuri, Paichayontvijit and Shen (2013) found that individual females were more trustworthy than individual males, but there were no significant differences in the trusting behaviours of males and females. Also, gender differences tended to be eliminated over time and if groups (same sex or mixed), rather than individuals, played the game. Also, Bonein and Serra (2009) found that, when the gender of the other player was not known, there were no gender differences in trusting and trustworthiness, while when the gender was known, there were no differences in trusting, but there were significant differences in trustworthiness — specifically, participants returned more money to same-sex partners, suggesting a gender pairing bias in trustworthiness. On the contrary, in Slonim and Guillen (2010), participants trusted game partners of the opposite sex more. Furthermore, Chaudhuri and Gangadharan (2007) found that males trust more than females in trust games and dictator games, but they found no gender differences in reciprocal behaviour. Taken together, these results suggest that any effects of voice pitch on trustworthiness attributions are due to relative differences within each of the sexes, rather than a single continuum that crosses gender.

Finally, voice quality — variations of which distinguish, for example, breathy, creaky and modal phonation (Klatt & Klatt, 1990) — might convey information about personality characteristics as well. For example, Laver (1968) suggests that harsh voices are correlated with more aggressive and dominant characteristics, while breathy voices are related to more submissive personalities, and Blood et al. (1979) found that both hypernasal and breathy voices — which can be symptoms of voice disorders — were rated more negatively than modal voices. A breathy phonation mode in female voices can also related to perceived femininity and attractiveness, in what has been called the “Sirenic Code” (Gussenhoven, 2016). Breathy male voices were also rated positively by female speakers (Xu et al., 2013), perhaps because breathiness in male voices can signal a reduction in aggressiveness.

These vocal characteristics, along with many others, can also be used to convey emotions in the voice. However, there seems to be no direct mapping between individual vocal features and individual emotions. For instance, f0 has been reported to increase for joy, but also for anger and fear, and to decrease for sadness and boredom; or, f0 range has also been reported to increase for joy, anger and fear (see review in Kappas, Hess & Scherer, 1991). In general, previous studies come to very different conclusions as to which vocal feature is a correlate of which emotion (for a review, see Table 23.2 in Scherer, Johnstone & Klasmeyer, 2003). Nevertheless, it is clear that people can recognise emotions from the voice alone (e.g. Banse & Scherer, 1996; Frick, 1985). In fact, there is consistency in vocal emotion recognition across very different cultures (e.g. Bryant & Barrett, 2008; Sauter, Eisner, Ekman & Scott, 2010), suggesting that vocal emotion expression might be universal (Cosmides, 1983). There are also links between emotional expression and trustworthiness (Boone & Buck, 2003). For example, trustworthy faces who expressed happiness were perceived as happier than untrustworthy faces, and untrustworthy faces who expressed anger were perceived as angrier than trustworthy faces (Oosterhof & Todorov, 2009). Also, happiness and gratitude were found to increase trust, while anger decreased trust (Dunn & Schweitzer, 2005). Since both emotions and trustworthiness are recognisable from the voice, it is reasonable to think that a similar interaction will be found in the auditory domain as well. Furthermore, because of the “halo effect” outlined earlier, vocal emotional expressivity is likely to influence personality judgments. Thus, we might expect that people exhibiting positive emotions will be attributed other positive traits, such as trustworthiness, too.

2.2.4 Personality attributions to machines

With an increased number of artificial agents operating within our social space, we might need to consider how to interact with them. As previously discussed, societies need trust to function, so it is likely that human-machine societies will need trust to function as well. In particular, what is needed is a way of inferring a machine's trustworthiness. But, in order to be perceived as trustworthy, a machine needs to be perceived as a social entity first, since trust is built in interaction.

Interacting with machines as if they were social agents does not seem to be a problem, though. In their seminal paper, Nass, Steuer and Tauber (1994) demonstrated that people unconsciously apply to human-machine interaction the same social rules that they apply in human-human interactions. This phenomenon has been referred to as CASA (Computers As Social Actors). In their first experiment, they showed that participants applied norms of politeness to their interaction with a computer, because they evaluated a computer's performance better when asked directly by that computer, than when asked by another computer. In the second and third experiments, participants distinguished different agents based on their voice alone, even though the voices were played on the same computer. In the fourth experiment, participants applied gender stereotypes to computers: praises from computers with a female voice were judged less convincing that praises from computers with a male voice. Finally, in their fifth experiment, they showed that participants believed they were interacting with a computer agent, rather than a human programmer. Importantly, all the participants in these experiments were experienced computer users, and, when asked directly, they stated that computers should not be treated as social agents. Corroborating evidence comes from studies of interactions with navigator systems (Large et al., 2017) or robots (Lee, Peng, Jin & Yan, 2006). Thus, people might unknowingly treat each type of interaction, including with computers, as essentially social, and will behave accordingly.

There is also evidence that personality traits are attributed to artificial agents. For example, Nass et al. (1995) showed that even approximate representations of personality traits in computer agents were recognised, and acted accordingly upon, by participants. In a following experiment, Nass and Lee (2001) added that the same holds for personality cues represented in a synthetic computer voice. Participants accurately recognised personality traits from the voices, and showed a preference towards a computer personality that matched their own in their behaviour and explicit evaluation of the agents. Several studies have also shown that personality traits are ascribed to robots as well (e.g. Syrdal, Dautenhahn, Woods, Walters & Koay, 2007; Hwang, Park & Hwang, 2013; Salem & Dautenhahn, 2015). Going back to the example opening this chapter, there is indeed evidence that navigator systems are also treated as social agents (Large et al., 2017). For example, participants in a simulated driving system interacted with their navigator system, applying the same conversational and social norms they would use in human interaction: politeness, turn-taking, back-channelling, fillers and hesitation, were all used by participants when conversing with the navigator system (Large et al., 2017).

While people perceive other people as trustworthy by default (G. R. Jones & George, 1998; Barclay, 2008), the same cannot be said for agents, especially in high-risk contexts such as factory work, where a misplaced trust in a robotic arm could cost a human worker's life. Thus, it is up to the manufacturers to make sure that the machines are safe to work with, but it is also up to the designers to design trustworthy-seeming machines. Since many of the machines that collaborate with us are disembodied (e.g. a navigator system or a mobile personal assistant), their voice should also be designed as to elicit trust. In the case of robots, which do have a body, the voice design should be an integrative part of the body design. For example, should big robots have deeper voices than small robots? Should human-like robots have natural human voices? Should machine voices in general have accents? Preliminary research from Andrist, Ziadee, Boukaram, Mutlu and Sakr (2015) on the Arabic language shows that accent and context interact in robots as well: participants believed that robots with a regional accent were more credible when they showed they were knowledgeable, whereas robots with a standard accent were more credible when they had little knowledge. A similar interaction is likely in other languages as well, although so far there is no data to prove it.

2.3 Summary

Trust is an essential aspect of social interactions, and it can be considered as two strongly related processes: the trusting behaviour of an agent, the trustor, and the trustworthiness of another agent, the trustee. Both these aspects are generally measured attitudinally, for example through personality questionnaires. However, as mentioned in the introduction (Section 1.2), such measures are not indicative of actual behaviour, and do not inform us about the development of trust.

Impressions of trustworthiness can also be evinced from the voice, and several studies looking at different aspects of voices have been examined. However, there is no conclusive argument as to which vocal features are perceived as trustworthy. One reason for this confusion is that different studies have used different methodologies, and many have looked at different expressions of trust, for example charisma, persuasion, deception, etc.

Furthermore, studies on trustworthiness attributions generally do not take context into account. However, this aspect is particularly important for Human-Machine Interaction: machines do not always work properly or as expected, so it is necessary to examine users' reactions both when the interaction goes smoothly and when it does not. In particular, the role that the machine's voice has in different contexts has not been investigated.

Framed in the literature discussed above, this thesis contributes to research on trust and speech science in several ways:

- It contributes to knowledge on characteristics of the trustee that elicit trusting behaviours. In particular, it concentrates on the trustee's voice;
- By modifying the trustee's behaviour, it provides new insights into trust development towards trustworthy and untrustworthy trustees;
- By using an implicit measure, it provides an actual representation of people's behaviour with different agents, and in different contexts.

3 Methodology

All the experiments presented in this thesis share the "repeated investment game" as part of their methodology. This method allows to implicitly measure trust in the form of monetary investments, and was originally devised by Berg et al. (1995).

There are two different ways to collect personality attributions: implicit or explicit measures. As previously seen (Section 1.2), some implicit measures are actual representations of behaviour (Hofmann, Gawronski, Gschwendner, Le & Schmitt, 2005); typical examples of implicit measures are response time, interpersonal distance, results obtained from games, etc. Implicit measures can provide evidence of cognition, inclination or attitudes which participants are not aware of. In fact, the outcome of implicit measures can become an index of the attitudes and cognition participants have no control or conscious access to. On the other hand, explicit measures are representations of someone's subjective beliefs about a situation which can be retrieved from qualitative or quantitative verbal data, such as opinions, ratings, questionnaires, etc. (Schwarz, 1999). Also, they are generally self-reflections on a past situation, but they do not usually measure current states (Greenwald, 1990). To give an example of the difference between these two types of data in a previously mentioned study, in Nass et al. (1994)'s experiments on the "Computers As Social Actors" paradigm, the participants explicitly stated that people should not behave socially with computers, yet they implicitly behaved so themselves. To give an example with trust, an explicit method would be asking participants to rate "How trustworthy does this voice sound" on a certain Likert scale. However, as previously mentioned (Section 1.2), subjective beliefs do not correlate with actual behaviour (De Houwer, 2006; Greenwald & Banaji, 1995; Wicker, 1969), and they only collect impressions at a certain time. Implicit measures are also influenced by preferences (Fehr, 2009a). If participants are given a questionnaire to fill in about the perceived trustworthiness of a certain agent, the response will generally reflect the participants' beliefs about what their first impressions are. Such first impressions are important, because they form the basis upon which the perceived impression is going to be developed, but they do not inform us about the attribution development process, which takes additional information such as behaviour, experience and attitudes into account. The investment game allows to account for this dynamic process as well.

3.1 Game theory

The investment game, as devised by Berg et al. (1995), is derived from game theory, a very influential economic theory that postulates how agents behave in interaction (Sanfey, 2007). A very famous example of game theory is the “Prisoner's Dilemma” paradigm. In the Prisoner's Dilemma, two criminals are arrested and imprisoned, in two separate cells, so that they cannot communicate with each other. The police does not have enough evidence to convict them, so they offer a deal to each criminal, individually: they can either testify that the other committed the crime, or they can remain silent. Thus, if criminal A testifies against criminal B, and criminal B remains silent, A will be set free and B will be sentenced to 3 years (or vice versa). If A and B both testify against each other, they will both be sentenced to 2 years. Finally, if both A and B remain silent, they will both be sentenced to 1 year (Poundstone, 1992). Testifying against the other player is an act of defection, while remaining silent is an act of cooperation. The best outcome for each player is when both cooperate, but no matter what the other player chooses, a player always earns more by defecting than by cooperating. Thus, assuming that the other player reasons in the same way, one player can be certain that the other player will defect, so the only rational choice is to defect as well. This theorem has greatly influenced the international political actions during the Cold War. The postulation of the perfect Nash equilibrium, i.e. the situation in which no player has anything to gain by changing his strategy, awarded Nash, Harsanyi and Selten the Nobel Prize in Economics in 1994.

The original Prisoner's Dilemma is one-shot, which it to say, each agent makes one action, and after that the game is finished. However, more relevant examples for this thesis are iterated games, where the agents have to make several actions. In the iterated Prisoner's Dilemma, players make the simultaneous choice of whether to cooperate or defect over an indefinite number of rounds. While the best strategy in one-shot Prisoner's Dilemma is to defect, the best strategy in the iterated version is the so called "tit-for-tat": defect if the other player defected in the previous round, cooperate if the other player cooperated (Poundstone, 1992). Of course, in case the number of rounds is known to the players, the optimal strategy is tit-for-tat until the last round, at which point both players should defect.

Another well-known example from game theory is the "dictator game". Player A, the dictator, is given an amount of money and can offer a share of it to player B. If player B accepts the offer, both players keep what they have. If player B refuses however, no-one keeps any money. Thus, the best strategy for player B is to accept any offer, since it will still be more than his/her initial endowment (which is 0). Assuming this, the best strategy for player A is to offer the minimum amount offerable. In this game, player B's decision is a punishment decision, not a cooperation one, and, by punishing, both players are worse off (Rand, 2016). In a repeated ultimatum game, Rand (2016) proposes that the best strategy for player B is to reject unfair offers, in order to induce Player A to make fairer offers in the next rounds. Again, in case the number of rounds is known to the players, the optimal strategy is this threat-bargain until the last round, where Player A should offer the minimum amount and Player B should accept it.

Generally speaking, game theory is the study of interactions between agents striving to achieve an optimal solution. Its applications span the disciplines of Economics and Political Science, and have been widely researched in Philosophy, Psychology, Biology, Anthropology, Zoology, Sociology, etc. Agents can be two countries at war, but also two children deciding how to split a cake or two romantic partners deciding what to do in the evening. While the first two examples are so called "zero-sum" games, where the total of gains and losses equals to zero, the last example represents a non-zero-sum game, where the winnings and losses of all players do not add up to zero. The wife might want to go to the cinema and the husband to the football match, but each of them would rather spend the evening with the other, than alone. Non-zero-sum games model cooperation, rather than conflict, because in this type of game individuals have more incentives to cooperate. Thus, while zero-sum games need to have a winner and a loser, non-zero-sum games are not so clear-cut, and there is space for cooperation and mutual benefit, if all the parties are trustworthy: this is were trust comes into play.

The investment game, as devised by Berg et al. (1995), is a non-zero-sum game. Participants in the role of player A are given $10, and can decide how much of this money to send to player B. This amount is tripled, so that player B can now decide how much of this tripled amount to send to player A. If A decides to send any money at all, B will have earned something, since he/she started with $0. If player B returns more than what A sent, player A will have earned something. If player B returns less, player A will have lost something. However, the sum of the winnings and losses does not equal to zero: even in the case in which player A sends nothing, player B will have earned $0, but player A will still have $10. The amount of money that player A sends is an implicit measure of his/her trust (Berg et al., 1995; Wout & Sanfey, 2008; Camerer, 2011); the amount of money that player B returns is an implicit measure of his/her trustworthiness. The investment game has been used to measure trusting behaviours in several studies since its formulation. A summary of such studies is given in the next section.

3.2 Uses of the investment game

Researchers have used the investment game to study trusting behaviours towards a range of characteristics. For example, several studies have used it to examine trustworthiness attributions and trusting behaviours for different genders (e.g. Boenin & Serra, 2009; Buchan, Croson & Solnick, 2008; Chaudhuri et al., 2013; Charness & Gneezy, 2012). Also, the trustworthiness of facial expressions, such as smiles, has been studied using the investment game (Krumhuber et al., 2007; Scharlemann, Eckel, Kacelnik & Wilson, 2001; Tortosa, Lupiáñez & Ruz, 2013). Past studies have also used the investment game to look at trusting differences in various countries (Willinger, Keser, Lohmann & Usunier, 2003; Croson & Buchan, 1999). There are no studies, however, that used the investment game to study trusting behaviours to voices. The only two studies to go in this direction are O'Connor and Barclay (2017) and Montano, Tigue, Isenstein, Barclay and Feinberg (2017), who looked at trust attributions to voices with different pitch. However, they used a one-shot game, which means they could not draw any conclusion about attitude development. Also, they asked participants to imagine the game, rather than actually play it, thus effectively nullifying the benefits of implicit measures.

Berg et al. (1995)'s investment game has also been modified to suit experimenters' needs. For example, Krumhuber et al. (2007) used a version of the game in which the players cannot decide exactly how much money they want to send or return, but they have a binary choice: the trustor has £5 and can either keep it or invest all of it; the invested amount is then doubled and the trustee can either keep it all or return £7.5 and keep £2.5. Similarly, in Tortosa et al. (2013), trustors start with €1, and can either keep it or invest it; if they invest it, it is quintupled, and the other player can either keep the entire amount or return half of it. Or, in Stanley et al. (2012), trustors could only invest in increments of $2 of their endowment, and in Slonim and Guillen (2010) they could invest in increments of $3. In most studies, however, trustors can decide the amount of money they want to invest in a continuous way (e.g. Bereczkei et al., 2013; Chaudhuri et al., 2013; Bonein & Serra, 2009; Chaudhuri & Gangadharan, 2007; King-Casas et al., 2005). Also, in experimental procedures, generally participants play the role of the trustor, although occasional studies ask participants to play the trustee (e.g. Bereczkei et al., 2013; Jordan, Hoffman, Nowak & Rand, 2016).

The vast majority of studies on implicit measures of trust used one-shot investment games. Thus, the collected data from these studies provides an implicit measure of trust as a first impression, but no inference can be drawn about the impression development process. A notable exception is Samson and Kostyszyn (2015), where participants played 10 rounds, and the behaviour of the trustee was controlled so that it returned 150% (± 10%) in the first 3 rounds, 0% in the 4th round and 150% (± 10%) again in rounds 5–10. Their intention was to examine whether people are impulsive (i.e., they react on what they were returned in the immediately preceding round) or follow a strategy (i.e., they have a consistent pattern of investments which considers the full history of the game), so they did not focus on the process of trust development. Nevertheless, this is the only study that explicitly manipulated the behaviour of the trustee over time.

From this review, it seems that variants of the investment game generally differ on a few aspects from the original: whether the investment choice is dichotomous or continuous; the percentage of multiplication of the invested amount; the variable of choice (trustor or trustee); the number of iterations. Although the investment game has been used to study trust towards a variety of trustees, no studies so far have used it to measure trust towards voices. The questions, then, remain: do voice characteristics influence trustworthiness attributions? And, is this influence constant over time, or does the effect of voices fade as participants gain experience of the trustor's behaviour? To answer these questions, the experiments presented in this thesis use an iterated investment game, where the participants play the role of the trustor, and the behaviour of the trustee, which is an agent speaking with different voices, is altered and controlled for.

3.3 Description of the investment game used here

In the experiments presented here, participants are given an amount of virtual money to start with. Before they make the first investment decision, they interact with the trustee, by hearing its voice. The participant then makes the first investment decision. This first investment is a representation of his/her first impression on the trustworthiness of the speaker at zero acquaintance. The money the participant invested is then multiplied, so that the trustee receives 3 times the invested amount. The trustee then returns an amount between 0 and the tripled investment, and the full transaction is shown to the participant on a screen. This sequence of actions is one round of the game. The participant receives a new endowment at the beginning of each round, and a game consists of 20 rounds. Participants do not know how many rounds they will play.

The interaction between the participant and the trustee, happening just before the participant's investment, consists of a pre-recorded sentence from a speaker, that is played to the participant. The voice used reflects the focus of the specific experiment. In Experiment 1, they hear either an SSBE- or a Liverpool-accented voice, in Experiment 2, either an SSBE-, London-, Plymouth- or Birmingham-accented voice, and so on. The trustee, in these experiments, is presented to the participants as a virtual agent in Experiments 1-4. Essentially, the trustee is a computer, whose behaviour is pre-programmed. In Experiment 5, the trustee is a Nao robot, whose behaviour is also pre-programmed, as described below.

The monetary returns of the trustee are pre-defined, so that the trustee simulates either a trustworthy behaviour (*generous* condition) or an untrustworthy behaviour (*mean* condition). In the *generous* condition, the trustee returns more money than the participants invested; in the *mean* condition, it returns less. Specifically, it returns 40% to 80% of the tripled amount in the *generous* condition in Experiments 1-4, and 0% to 40% of the invested amount in the *mean* condition in Experiments 1-4. A visual representation of one round of the investment game is shown in Figure 1.

The exact amount returned at each round is fixed between games, as shown in Table 1. As it is possible to see, there are a few rounds in the *mean* condition in which the trustee is actually returning more money than the participant invested (these are the rounds with 120% returns of investments). There is a reason behind this. First, if the trustees were always returning the same percentage of investments (e.g. always 70% in the *generous* condition and 30% in the *mean* condition), participants might have become aware of this manipulation after a few rounds, and might have started investing the same amount at each round. Thus, a range of returns, rather than a dichotomous behaviour, was needed. This range also needed to be wide enough to keep participants engaged in the game, so a range spanning 50% was chosen. Since the range needed to be the same in both behaviour conditions,

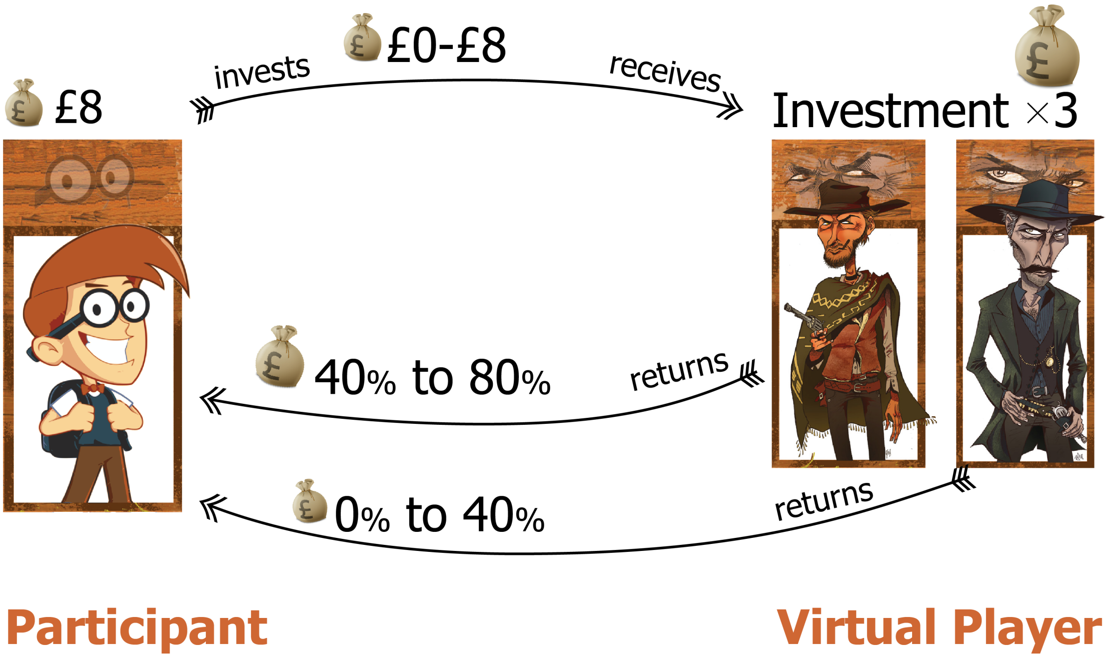


Figure 1: Scheme of one round of the investment game. The "nice guy" is the *generous* trustee, the "bad guy" is the *mean* one. Image credits: © Zach Bellissimo 2008 and Frank Loesche.

this is how the two behaviours came to be. Of course, it is possible that the few rounds in which the virtual agent trustees were returning slightly more to the participants did not go unnoticed, and that participants' reactions were affected by this. Thus, in order to examine a fully untrustworthy behaviour as well, the behaviours were slightly modified in Experiment 5, where the percentages of return were 50% to 80% in the *generous* condition, and 0% to 30% in the *mean* condition. In this way, it was possible to examine the patterns of investments in a condition where the trustee always returned less than the participant invested, while still keeping a fixed, albeit narrowed, range. The return patterns for Experiment 5 are reported in Appendix B.

Round	Generous	Mean	Sentence block 'YZ'
1	150%	30%	Welcome to the investment game. I hope we will enjoy playing it
2	150%	30%	In my opinion, we should always invest in one another
3	180%	60%	I think we can do better. I promise that I will not let you down

Round	Generous	Mean	Sentence block 'YZ'
4	120%	0%	If we both invest in each other, we will surely raise our earnings
5	180%	60%	We could finish the game better off than this, if only we tried harder
6	210%	90%	You have to trust that I'm going to cooperate until the last round
7	120%	0%	We can both win the game, but we have to keep sharing our money
8	120%	0%	I trust you, and I am sure that we can both benefit from each other
9	210%	90%	I am not a greedy person, and I believe we should share these earnings
10	210%	90%	I am not a greedy person, and I believe we should share these earnings
11	120%	0%	I will return more of your investments, you have to trust me in this
12	150%	30%	When the game ends, I promise that we will both be satisfied with the outcome
13	240%	120%	Remember this: it's not convenient for me to keep all your investments
14	180%	60%	No matter the number of rounds, we should trust each other until the end
15	240%	120%	There is no better tactic than to keep investing and returning
16	210%	90%	I will show you that co-operation is the best option for us
17	240%	120%	Let's keep sharing, and our earnings will grow much bigger than they are now
18	180%	60%	I promise that I am going to return more money from now on
19	240%	120%	I will always return, because there's no point in me doing otherwise
20	150%	30%	If we want to see our funds growing, we have to share until the end

Table 1: Returns of investments in Experiments 1–4, and samples from one block of sentences. Note: the investment is tripled, so the return percentage is tripled to show the return from the original investment.

Table 1 also shows one of the blocks of sentences that were devised for the game. Participants played either 2 games (in Experiments 1, 3, 4, 5), or 4 (in Experiment 2). Each of the speakers who were recorded to provide the stimuli for each experiment read a 20-

sentence “block” for each game that the participants would play; thus, there were 2 blocks of 20 sentences in Experiments 1, 3, 4, 5 and 4 blocks in Experiment 2. Thus, 80 different sentences were devised, so that even participants playing 4 games would always hear different sentences. Each of the 4 resulting blocks of 20 sentences was given a code: “AB”, “CD”, “WX”, “YZ”. While the first sentence of each block was semantically neutral (e.g. “Hello, let’s get started with the investment game”), all the others were about strategies to follow in the game (e.g. “We have to help each other out, it’s the only way to win the game”). The full set of sentences that all the speakers recorded for the experiments presented here read can be found in Appendix A. While it could be argued that such invitations to be trusting might have influenced participants’ behaviour, specifically to be more trusting than they would have been otherwise, there are a few reasons to believe that this was not the case. First, the content of the sentences was completely unrelated to what was actually happening in the game, and results from all the experiments consistently show that participants trusted trustworthy virtual partners and did not trust untrustworthy virtual partners. Also, theoretical work on trust states that trust cannot be forced (Gambetta, 1988; Zak, 2003), and that imperatives like “trust me” do not work unless trust is present in the first place (Boon & Holmes, 1991). All the sentences, apart from the first one of each block, were encouraging trusting behaviours. For this reason, it could also be argued that this could have caused participants to trust more when the virtual partner was actually trustworthy, and to trust less when the virtual partner was untrustworthy, because of the incongruity between the virtual partner’s behaviour and its words. In real life encounters, however, untrustworthy individuals will not openly assert their untrustworthiness, since it is this deception that makes it advantageous to be untrustworthy in the first place (Boone & Buck, 2003). Thus, these sentences constitute an ecologically valid contribution to the simulation of trustworthy and untrustworthy partners.

3.4 Critiques of the investment game

There have been some critiques to the use of the investment game as a measure of trust. Some researchers have raised the concern that participants' investments might be influenced by their individual risk attitudes and behaviour, and that trusting might be confounded with risk-taking (Fehr, 2009a; Karlan, 2005). Contrasting this, several studies found that risk attitudes did not predict decisions in the investment game, and concluded that risk attitudes affect risky decisions but not trusting decisions (Houser, Schunk & Winter, 2010; Ashraf, Bohnet & Piankov, 2003; Eckel & Wilson, 2004; Fetchenhauer & Dunning, 2012). The difference between trusting decisions and risky decision is that the success of the former is attributed to another agent, while the success of the latter is attributed to a statistical percentage, which is known or at least assumed by the person who takes the risk. Thus, for example, trust games are played between two agents, while risk games are played between an agent and a computer which returns a random amount of money based on a certain distribution, but it is not given agency (Kosfeld, Heinrichs, Zak, Fischbacher & Fehr, 2005). Also, in their survey of trust measures Ben-Ner and Halldorsson (2010) mentions that risk attitudes are not related to trusting or trustworthiness. From the point of view of the trustor, people felt "betrayal aversion" only when they were deceived in a trust game, but not when they lost in a gamble (Fetchenhauer & Dunning, 2012). People who lose on a gamble just feel they had bad luck, while people who trusted a game partner, who then defected them, feel they had been exploited (Bohnet & Zeckhauser, 2004). Thus, people might be more willing to take a risk with a certain probability of losing than to trust when facing the same probability of being deceived (Fehr, 2009a). These conclusions are also supported by a study by Kosfeld et al. (2005), who found that oxytocin, a hormone that influences prosocial behaviour, affects trust decisions in an investment game, but not risky decisions in a risk game. Additionally, Reuter et al. (2009) found that people who have a variant of the oxytocin receptor gene, which is believed to increase trust, trusted more in a trust game, but did not risk more in a risk game. McCabe, Houser, Ryan, Smith and Trouard (2001) also found that regions of prefrontal cortex were more active when participants in a trust game were playing with another human than with a computer known to be following an algorithmic behaviour. This suggests that decisions in the game go beyond reinforcement learning to encompass social relationships. Thus, behavioural and biological data suggest that trusting and risky decisions might be different phenomena which only partially overlap; as such, the investment game can be used as an accurate representation of trusting behaviour.

3.5 Statistical and experimental procedures

The main statistical tool used in the experiments presented here is the mixed-effects linear model, which is sometimes also referred to as multilevel analysis. This technique adds to linear regression techniques by allowing to specify random effects, as opposed to fixed effects in the models. The latter are supposed to represent systematic changes to the dependent variables, while the former represent random variability which is unaffected by the independent variables. For example, if one wishes to study what affects athletes' performance in a running competition, elements such as body weight and height, age, type of clothing and shoes, amount of sugars ingested prior to the competition, etc. would all be fixed effects, because they are all assumed to affect the runners in a linear fashion. On the other hand, some of the runners might have had a troubled sleep due to agitation before the performance; others might have had a fight with their partner which might distract them during the event; and so on. All these factors are thought to be random, because they typically are not controlled for. By specifying that the speed of each runner is a random factor, the model will allow for idiosyncratic differences in the intercept or slope of each individual runner, which would otherwise result in inflated variability in the data. Mixed-effects models are sometimes preferred to ANOVAs, for example, because they allow to discard the by-item, by-subject problem (Baayen, Davidson & Bates, 2008).

The models built in the experiments presented in this thesis are rather complicated, because all have several predictors. For this reason, the models were built using forward stepwise selection: models containing one predictor each are compared to a null model with a likelihood ratio test (anova() function in R). The model with the lowest AIC (Akaike Information Criterion, see Aho, Derryberry & Peterson, 2014) is then selected as the next null model to which new models with new predictors are compared. This is repeated until all possible combinations of predictors and interactions of predictors (up to 3-way) have been tested. The effect structure in the models is as follows: participant id is a random effect, for the reasons specified above; sentence blocks are also random effects, since they can be thought of as a random sample of all sentences that players in an investment game could use to prompt their partners to trust them. Participants' monetary investments are the dependent measure. All the manipulations of the game, such as behaviour of the trustee, trial of the game (referred to as "game turn"), accent, pitch, etc. are fixed effects.

All statistical analyses were carried out with the R software (R-v3.1 to R-v3.3, see R Core Team, 2016) in the RStudio GUI. Mixed-effects models were fitted using the lme4 package (Bates, Mächler, Bolker & Walker, 2015); data were plotted using the ggplot2 package (Wickham, 2009); tables were created using the stargazer and kable functions. Regarding experimental procedures, unless otherwise stated, experiments were scripted and presented to participants with E-Prime~2.0. Data were logged in E-Prime, opened in E-Merge and exported to `.csv` in a post-processing step. All resulting `.csv` files were then prepared and analysed in R.

4 Experiment 1 — Accents

4.1 Introduction

We make initial judgements about personality based on very limited evidence, such as seeing a face for 100ms, or hearing the word "hello" (Willis & Todorov, 2006; McAleer et al., 2014). Many physical and behavioural characteristics (e.g., facial appearance and expression, dress, voice quality, accent) contribute to the formation of these impressions, as was previously discussed. It is still unclear, however, what specific vocal characteristics contribute to attributions of personality traits.

In particular, accent differences can suggest personality stereotypes, even to non-native speakers (Ladegaard, 1998; Bayard et al., 2001), and native accents may be perceived as more trustworthy than non-native accents (Lev-Ari & Keysar, 2010). Accents, though, are often intrinsically related to geographic regions (with the exception of accents associated with social class, such as Standard Southern British English – SSBE), and stereotypes based on general socio-economic perceptions of particular regions may impact on personality attributions. Some studies have shown that standard accents such as SSBE are rated as more pleasant and attractive than, for example, city accents such as Liverpool or Birmingham (Bishop et al., 2005; Dixon et al., 2002; Fuertes et al., 2012). People may not actually be very effective at localising accents however (Bayard et al., 2001), and it is possible that some vocal characteristics of accents may mediate trust judgements independent of regional stereotypes. Furthermore, most research on accent attributions focuses on immediate impressions, without taking into account how the attribution might evolve over time according to the speaker's behaviour.

The focus of the first experiment presented here is on trust attributions to different accents. Since data on implicit trustworthiness attributions to accents is scarce, this experiment was also meant as a baseline comparison with explicit data from previous sociolinguistics studies. Thus, two accents that have often been placed at opposite ends in trustworthiness questionnaires, SSBE and Liverpool English (Bishop et al., 2005; Dixon et al., 2002; Fuertes et al., 2012), were chosen as experimental manipulations. Plus, the experiment was designed so that each participant would play one game with an untrustworthy partner and another game with a trustworthy partner. As previously mentioned, a trustworthy partner was one who overall returned more money to the participant than the participant invested, while an untrustworthy partner overall returned less than the participant invested.

Apart from examining the contribution of accent as a main effect — i.e., do people implicitly trust an SSBE speaker more? — the focus of this experiment was also to study the interaction between the accent and the behaviour of the virtual players. In other words: given that the virtual player always returns either a generous amount of money or a poor amount of money, would there be any differences in the relative monetary investments to the SSBE-accented virtual player and the Liverpool-accented virtual player? Very little research has examined the effect of behaviour on accent attitudes, or voice in general. For example, in a study in Human-Robot Interaction, Andrist et al. (2015) manipulated the knowledge and rhetorical ability of robots acting as tour guides, as well as their accent (Modern Standard Arabic and Lebanese), and examined how much Lebanese participants accepted their suggestions. They found that participants complied more with the Modern Standard Arabic-accented robot when the robots had low knowledge, but preferred the Lebanese-accented robot when the robots expressed high knowledge and high rhetorical ability. Also, Giles, Williams, Mackie and Rosselli (1995), examining the “English only Movement” in California, found that Anglo-American speakers arguing against this movement were more persuasive than those in favour, and that Hispanic-American speakers speaking in favour of this movement were more persuasive than those against it. Also, Cargile (1997) found that attitudes towards Mandarin Chinese-accented speakers were different in the context of a job interview and of a college classroom. However, Cargile (1997)’s final argument that more studies on accent attitudes in context are needed, seems to have gone amiss, since not much more research on the topic had been conducted since then.

This chapter deals with the effect of accent on implicit judgments of trustworthiness, and with their interaction with behavioural cues. It presents results of an investment game where participants played with virtual agents which were associated with different accents and behaviour conditions. In particular, participants played with virtual agents which had either a Standard Southern British English or a Liverpool accent, which were behaving either *generously* or *meanly*. A summary of the experimental conditions can be found in Table 2.

Accent	Behaviour
SSBE	Generous
SSBE	Mean
Liverpool	Generous
Liverpool	Mean

Table 2: Experimental conditions of the investment game. Participants played one game with one accent and one behaviour condition, and one game with the other accent and behaviour condition, which were counterbalanced within participants.

Explicit trustworthiness data were also collected. At the end of the games that participants played, they were also asked to explicitly rate the speakers they heard for trustworthiness. The goal of these ratings was twofold: firstly, to verify to what extent the accented player's behaviour in the game would influence a subsequent explicit trustworthiness rating of the same voice; secondly, to understand how explicit ratings compare to actual behaviour in an investment game, and to previously reported explicit ratings in the existing literature.

4.2 Method

4.2.1 Participants

There were 44 native British English participants (35 females, 9 males) aged 18–45 (median = 19, SD = 6.8). They were university undergraduate students who received course credit for participation. The participants reported their places of origin, which were then clustered according to 5 of the 6 regions of the U.K. mentioned in Bishop et al. (2005): southwest England (n = 31), southeast England (n = 9), Midlands (n = 2), Scotland (n = 1) and Wales (n = 1).

4.2.2 Stimuli

Two male native British English speakers of SSBE and Liverpool accents were recorded in a sound-attenuated booth. Their voices were used as identifiers for the trustees in the investment game. In addition, 5 male native English speakers from Edinburgh, Birmingham, South London, Huddersfield and Bournemouth, and five male foreign English speakers from Germany (Saxony), France (Normandy), Italy (Tuscany), Greece (Macedonia) and India (National Capital Region) were also recorded. Utterances from all these 12 speakers were used in the questionnaire that was administered to the participants at the end of the game. Each speaker read two blocks of 20 sentences, each of which would be played at the beginning of each round of the game, described below. All the sentences had approximately the same length (mean number of syllables per sentence = 16.6, SD = 1.08), in order to ensure that participants were exposed to both speakers an equal amount of time. As previously described, apart from the first utterance of each block, which served for the virtual player to introduce himself, all other utterances were about strategies to follow in the game, e.g.: “I’m going to return more money now, if you invest more as well”; “Remember, there is potential for earning, if we both trust each other”; “The goal of the game is to earn as much money as possible”. Participants heard one block of utterances from one virtual player, and the other block from the other virtual player, in a random order. All recorded utterances were amplitude-normalized, and a noise-removal filter was applied.

4.2.3 Procedure

Participants played a repeated investment game, as explained in Section 3.3. Participants sat in a computer booth, wore over-ear headphones and carried out the whole experiment on a computer. They were told that the goal of the game was to earn as much money as possible, and that mutual co-operation with the other game partner would lead to greater profit. They were informed that they could not verbally interact with the other player, but that they would hear an utterance spoken by him at the beginning of each round. The participant played the role of the trustor, while the trustee was a computer program. The participant started with a notional sum of £8 at the beginning of each of the 20 rounds of a game. He/she then decided whether to invest all, part, or none of it with a virtual player, by pressing the corresponding number key. The trustee then received three times the amount that the participant invested. The trustee was programmed to have one of two behaviours, either returning 120% to 240% of the invested money to the participant (*generous* condition) or 0% to 120% (*mean* condition). The exact return pattern was fixed: the *generous* virtual players always returned 150% in the first round, 150% again in the second round, 180% in the third round, and so on, as explained in detail in Section 3.3. Thus, in the *generous* condition, if the participants invested any fraction of the money they were given, they would end the round with more money than they started with. The money that participants earned at each round was safely stored in a "bank". This was visible on the computer screen, as a reminder of how much money they were making in total, and how well they were faring in the game. Participants could not subsequently withdraw money from this "bank" however, but they could only make an investment out of the £8 they received at the beginning of each round. Participants heard an utterance at the beginning of each round, before they could make an investment decision.

Altogether, there were four accent-return conditions (Liverpool-generous, Liverpool-mean, SSBE-generous, SSBE-mean). Each participant engaged in two games, one for each accent, and one for each behaviour (*generous*/*mean*), with a different set of 20 sentences heard for each of the two games played. Each round of the game proceeded as follows: participants heard the utterance from the virtual player; they indicated how much of their £8 they wished to invest, in integers from 0 to 8 (by pressing a digit key); they saw a summary screen with all the monetary transactions to and from the virtual player that had happened during the round, including the return on their investment. After the end of the last round of the game, a screen transitioned the participants to a new game, with 20 more rounds. The virtual agent they played with in each game corresponded to one of the two accented speakers, SSBE or Liverpool.

After the participants finished the two games, they completed a questionnaire on the same computer display. Firstly, they were played two utterances from the first person they had played with, and were asked to rate how much they liked his voice, and to state why. Then, they were asked if they could recognise his accent. Then, they were asked to rate how much they agreed or disagreed with the following questions on the perceived honesty and sincerity of the virtual player (taken from Rau, Li & Li, 2009): “This person was sincere”“;”this person was interested in talking with me“”; “this person wanted me to trust him”“;”this person was honest in communicating with me“. The same four questions were then repeated after having played two utterances from the second virtual player. Then, they heard two utterances from each of the 12 speakers (including the ones whose first language was not English) in random order, and were asked to rate the speaker’s voice on a 7-point Likert scale (1 = very untrustworthy, 7 = very trustworthy). Finally, participants completed a short background questionnaire, where they were asked their age, gender, city of origin, and what accent they spoke. As a final question, participants were also asked what accent they would like a robot to have; this is discussed in full later, in Section 8.5. This data was kept anonymous, and was used to group participants into the 5 regions described above. The total duration of the experiment was approximately 15 minutes.

4.3 Results

4.3.1 Investment game

The overall investment pattern was dictated by the virtual player’s behaviour, with participants investing consistently more with the generous virtual player (Figure 2), regardless of his accent (Figure 3). A linear mixed-effects model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC (Akaike Information Criterion, Aho et al., 2014), with investment as dependent variable, accent, behaviour (*generous*/*mean*) and game turn (the 20 rounds of the game) as independent variables, and participants and sentence block as a random factors.

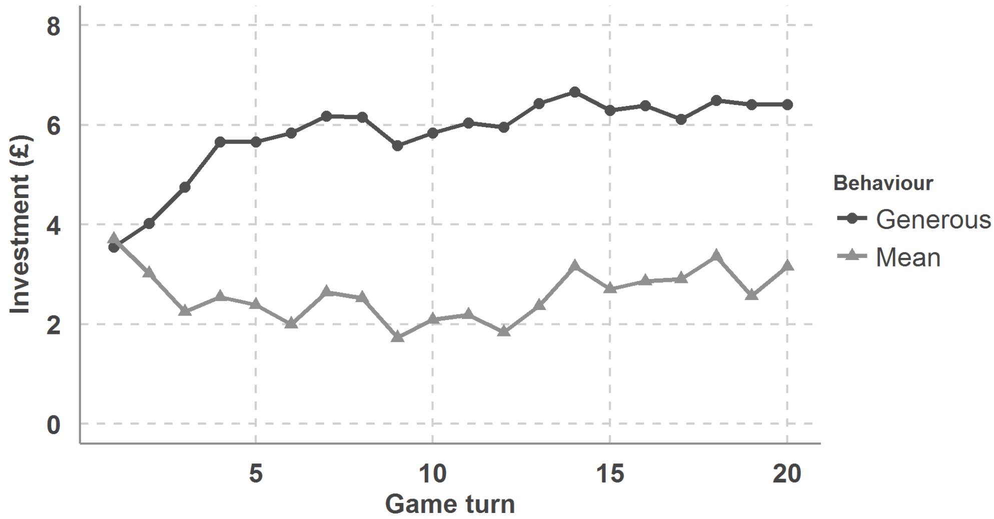


Figure 2: Average investments in the *generous* and *mean* conditions.

The model showed a main effect of behaviour ($\chi^2(1) = 984.25, p < .001$), with higher investments to the *generous* virtual players (average = £5.8) than to the *mean* virtual players (average = £2.6). There was also a main effect of game turn ($\chi^2(1) = 65.52, p < .001$), with higher investments in the second half of the game. There was also a significant interaction between behaviour and game turn ($\chi^2(1) = 40.80, p < .001$): as shown in Figure 2, investments increase almost linearly in the *generous* condition, while they decrease in the first half of the game, and increase in the second half of the game, in the *mean* condition. There was no main effect of accent ($\chi^2(1) = 0.51, p = .47$), but there was a significant interaction between accent and behaviour ($\chi^2(1) = 7.36, p = .007$), and a significant three-way interaction between accent, behaviour and game turn ($\chi^2(1) = 12.41, p < .001$). There was also A significant interaction between accent and game turn ($\chi^2(1) = 22.14, p < .001$): as can be seen from Figure 3, the difference of the overall investments between SSBE and Liverpool is positive in the first half of the game and negative in the second half of the game.

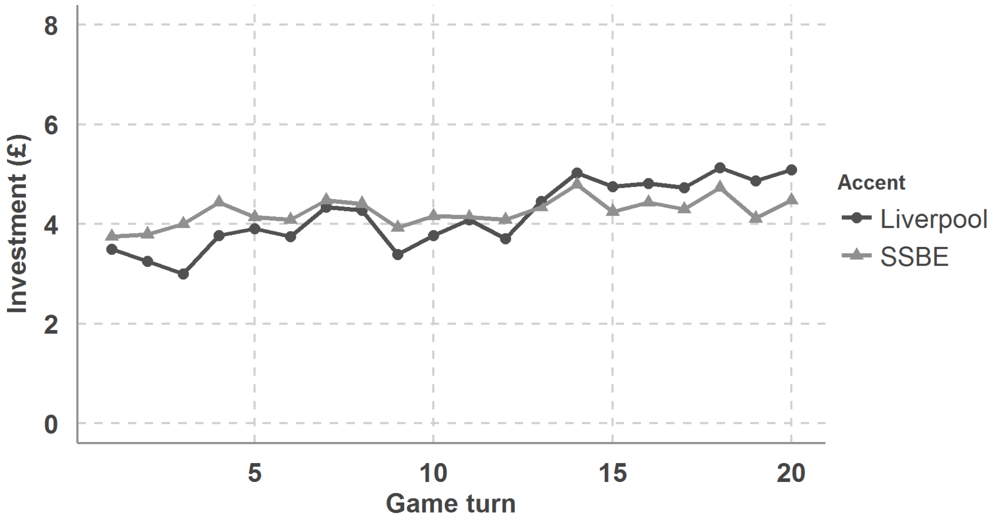


Figure 3: Average investments to the SSBE and Liverpool speakers

In order to examine the interactions, a post-hoc mixed-effects model with game turn and accent as predictors, and participant and sentence block as random factor was fitted to the data divided in the two behaviour conditions. In the *generous* condition, there was a main effect of accent ($\chi^2(1) = 4.93$, $p = .026$): as can be seen from Figure 4, participants overall invested more money with the SSBE-accented virtual players (average = £6.34) than the Liverpool-accented ones (average = £5.30). There was also a main effect of game turn ($\chi^2(1) = 151.77$, $p < 0.001$), with an overall increase in investment as the game proceeded, but no significant interaction between accent and game turn. In the *mean* condition, there was also a main effect of accent ($\chi^2(1) = 4.94$, $p = .026$), with higher overall investments with the Liverpool-accented virtual players (average = £3.06) than the SSBE-accented ones (average = £2.14). There was no effect of game turn, but there was a significant interaction between accent and game turn ($\chi^2(1) = 36.44$, $p < .001$): as can be seen from Figure 5, investments to the SSBE-accented virtual players decreased at the beginning of the game and then remained somewhat constant in the second half of the game, while investments to the Liverpool-accented virtual players increased in the second half of the game.

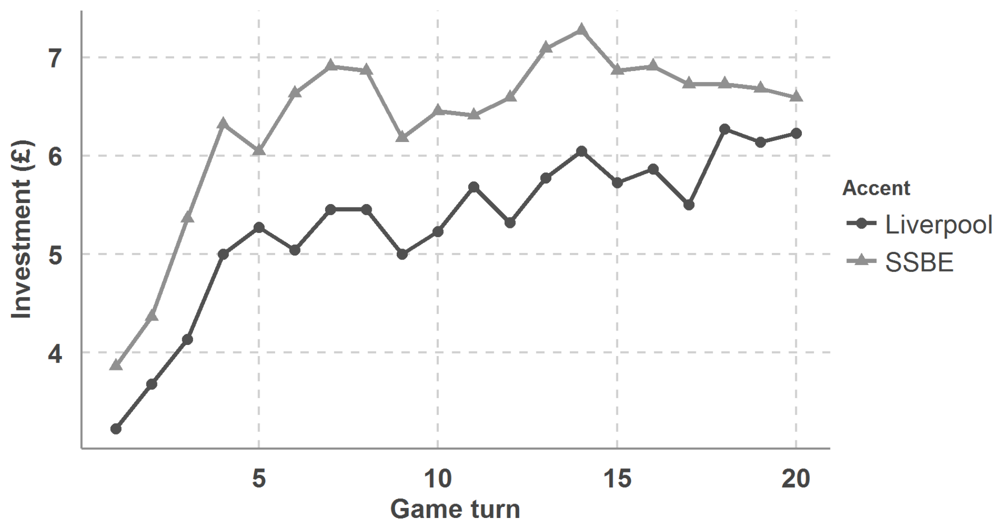


Figure 4: Average investments to the SSBE and Liverpool speakers in the *generous* condition.

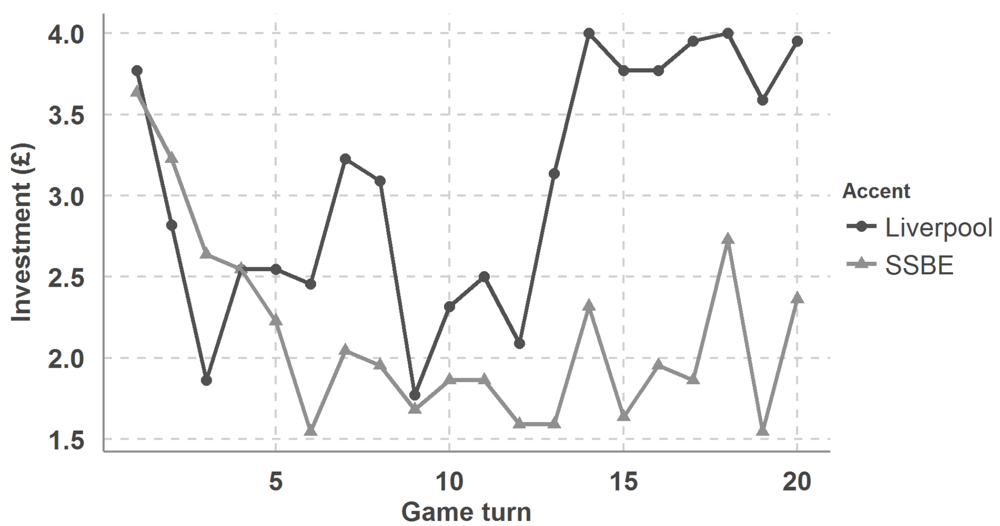


Figure 5: Average investments to the SSBE and Liverpool speakers in the *mean* condition.

Thus, differences between the responses to the accents emerge when considering the two behaviour conditions separately. When the virtual player was generous, participants consistently invested more with the SSBE-accented player throughout the game (Figure 4), supporting findings of relative trustworthiness of SSBE (Bishop et al., 2005; Fuertes et al., 2012). When the virtual player was mean, participants initially invested more with SSBE, but after three rounds the pattern reversed, and they subsequently invested more with the Liverpool-accented player (Figure 5), even though the pattern of investment return between accents was the same.

4.3.2 Questionnaires

Voice liking

Participants reported liking a voice more when it was associated with a *generous* behaviour (average = 4.93) than a *mean* behaviour (average = 2.97), independent of its accent. A one-way ANOVA revealed a significant effect of behaviour ($F(1, 85) = 84.05, p < .001$), but no effect of accent ($F(1, 85) = 3.07, p = .083$), and no interaction. Thus, speaker behaviour influenced liking ratings of that voice, but not accent.

Explicit honesty and sincerity evaluations

The 4 questions about the perceived honesty and sincerity of the speaker, taken from Rau et al. (2009), were the dependent measure in a mixed-effects model with behaviour and accent as predictors and participant as random factor. There was a main effect of behaviour ($\chi^2(1) = 156.82, p < .001$): participants overall gave higher ratings to the speakers in the *generous* condition than in the *mean* condition (average ratings = 5.65 and 3.19, respectively). This, together with the result that participants invested more money with the *generous* virtual players than the *mean* players, confirms that they understood how the virtual player was behaving and reacted accordingly. There was no effect of accent, and no interaction between behaviour and accent.

Trustworthiness questionnaire

A one-way ANOVA was performed on the trustworthiness ratings to the Liverpool and SSBE speakers. These speakers were the same that participants heard during the game. The ANOVA revealed that participants rated the speakers that had been associated with a *generous* behaviour as more trustworthy than the speakers that had been associated with a *mean* behaviour (average = 5.05 and 4.00, respectively, $F(1, 85) = 84.05$, $p < .001$). There was no effect of accent, and no interaction.

Regarding the trustworthiness rating of all the speakers, Figure 6 shows the mean trustworthiness ratings of the 12 sampled speakers. A one-way ANOVA on the trustworthiness rating of all speakers, apart from the Liverpool and SSBE ones heard in the game, revealed a main effect of accent ($F(9, 430) = 5.52$, $p < .001$). Post-hoc comparisons using the Tukey HSD test showed the following pairs to be significantly different: Birmingham was rated significantly more trustworthy than Huddersfield, French and Greek ($p < .001$, $p < .001$ and $p = .022$, respectively), Edinburgh was rated significantly more trustworthy than Huddersfield, French and Greek ($p < .005$, $p < .001$ and $p = .028$, respectively) and Indian was rated significantly more trustworthy than Huddersfield and French ($p = .022$ and $p < .005$, respectively). This pattern of ratings differed somewhat from previous trust attributions to accents of English (e.g. (Bishop et al., 2005; Fuertes et al., 2012)), where, for example, Birmingham accents scored relatively low and Yorkshire accents scored much higher. In this study, however, there was only a single speaker representing each region, and idiolectal characteristics clearly affected ratings in addition to any regional associations. Furthermore, in general speakers of English as a second language received lower ratings than native English speakers (average trustworthiness rating of L1 speakers = 4.22, of L2 speakers = 3.86, $t(437.99) = 2.61$, $p = .0095$), supporting previous findings (Frumkin, 2007; Fuertes et al., 2012; Lev-Ari & Keysar, 2010).

The ratings for the SSBE and Liverpool voices are likely to be influenced by participants' foregoing experiences in the investment game, and for this reason a direct comparison with the other speakers would be meaningless. Figure 7 shows the trustworthiness ratings of

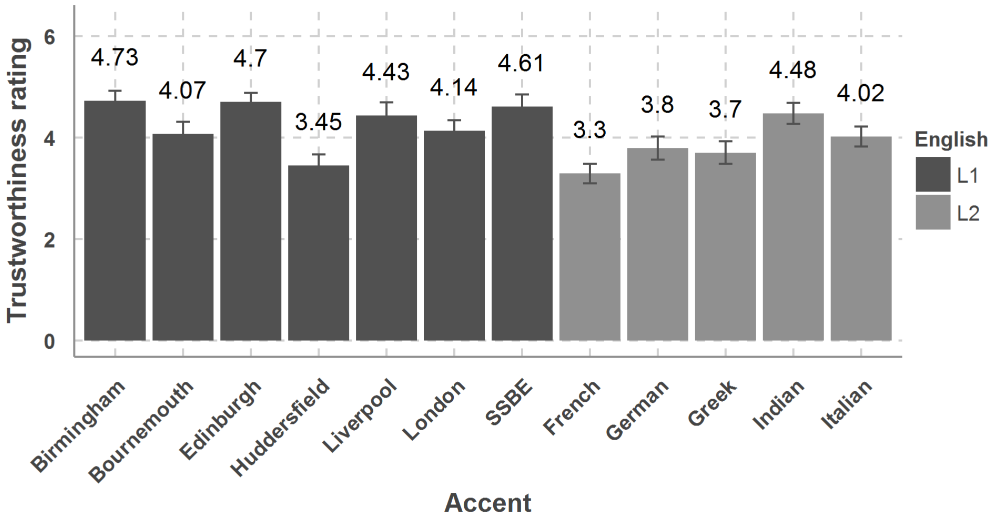


Figure 6: Trustworthiness rating of all speakers.

SSBE and Liverpool, grouped by the two behaviour conditions. Ratings in the Liverpool-generous were higher than in the Liverpool-mean condition ($t(42) = 2.51$, $p < .05$), whilst the difference between the SSBE-generous and SSBE-mean condition approached significance ($t(40) = 1.87$, $p = 0.07$). Thus, it appears that overall the explicit trust ratings are in line with participants' investment behaviour.

Accent identification

Of the 44~participants in the experiment, only~4 correctly identified the provenance of the Liverpool accent, and~11 the SSBE accent. Such a poor performance might be due to the fact participants were given a blank space to fill in with an accent label. It is possible that, had they been given a list of options to choose from, their performance might have been higher. Still, this poor performance indicates that regional stereotypes may play a minor role in trust attributions, compared to idiolectal characteristics. Furthermore, a post-hoc mixed-effects model adding participants' own accent of origin as a predictor of investments was not significant ($\chi^2(1) = 1.34$, $p = .25$). This result might not be particularly meaningful

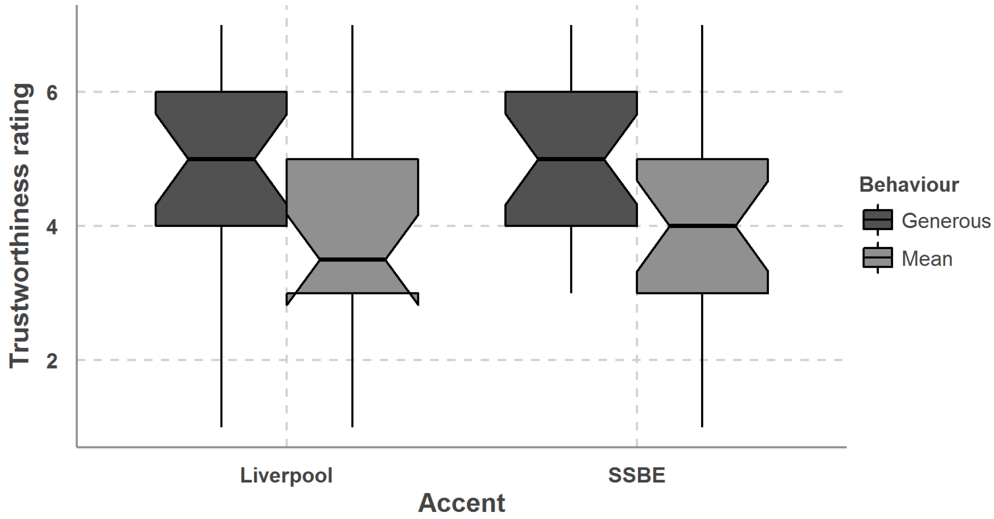


Figure 7: Boxplot showing the trustworthiness ratings of SSBE and Liverpool in the two behaviour conditions.

though, since the participant sample is not balanced in terms of their own accent. Thus, explicit social stereotypes and accent familiarity seem to play a minor role in implicit trust attributions, compared to idiolectal characteristics or stable phonetic differences between accents.

4.4 Discussion

Results from Experiment 1 clearly show that participants learn the behaviour of the virtual player fairly quickly, as indicated by the pattern of investments shown in Figure 2: participants invest more money with the *generous* virtual player and less with the *mean* virtual player. Furthermore, although no accent is overall trusted more than the other in the game, the interaction between accent and behaviour shows that participants trust more the SSBE-accented virtual player in the *generous* condition, and the Liverpool-accented virtual player in the *mean* condition.

In the *generous* condition, the higher investments with the SSBE speaker reinforce previous findings regarding preferences for standard accents over regional accents (e.g. Bishop et al., 2005; Dixon et al., 2002; Fuertes et al., 2012; Giles, 1970; A. M. Wilkinson, 1965). In line with "halo effect" theories (Nisbett & Wilson, 1977; Wetzel et al., 1981), someone who evaluates a standard-accented speaker positively, for example in terms of prestige, might implicitly extend this judgments to trustworthiness. Standard accents might also be clearer and more intelligbile (Floccia, Goslin, Girard & Konopczynski, 2006; Evans & Iverson, 2004), since most of the population will have listened to them in official radio or television programmes (Przedlacka, 2008). Although what the speakers said during the investment game was not related to what was actually happening in the game, participants are likely to have paid attention to it, and less intelligible utterances might have been penalised.

Interestingly, the accent of the speaker strongly affected how participants reacted to negative behaviour on the part of the virtual player. The speaker with the standard accent initially attracted higher investment, but showed a greater drop in investment than the regional speaker once the negative pattern of returns became evident. This result seems to indicate that socially prestigious accents may incur a more negative response to perceived unjust behaviour. Moreover, the investment patterns within the two behaviour conditions interestingly suggest that initial voice-based stereotypes (which may or may not be related to the actual speaker's accent, given the poor identification performance) are maintained and reinforced when the perceived behaviour is congruent with the impression that was initially formed (e.g. SSBE in the *generous* behaviour or Liverpool in the *mean* behaviour) but are reverted and penalized when the perceived behaviour is incongruent (e.g. SSBE in the *mean* behaviour).

This is in line with previous studies that found that prestigious accents were more appropriate in certain contexts, and regional or foreign accents in other contexts (Andrist et al., 2015; Giles et al., 1995). Experiment 1 is a further step in that direction, though, since not only does it provide evidence that accents might be trusted differently based on the behavioural context in which they are embedded, but it also suggest that such an "attitude" is implicit.

The virtual player's behaviour in the game also influenced post-game explicit rating of the same player's voice. Perceived trustworthiness, voice liking, and honesty and sincerity were all higher if the voice was associated with a *generous* behaviour in the game. This suggests that implicit impressions influence explicit ratings after the interaction has taken place.

4.5 Conclusion

In the *generous* condition, there was higher investment with the SSBE speaker, reinforcing previous findings of higher trustworthiness attributed to standard accents (e.g. Bishop et al., 2005). More interestingly, the accent of the speaker strongly affected how participants reacted to negative behaviour from the virtual player. The speaker with the non-regional (standard) accent initially attracted higher investment, but showed a greater drop in investment than the regional speaker once the negative pattern of returns became evident. This preliminary result intriguingly indicates that socially prestigious accents may incur a more negative response to perceived unjust behaviour. Data on accent attributions gathered with traditional sociolinguistic methods, such as questionnaires, would not have informed us about this temporal dynamics. Thus, it seems clear that other methodologies, such as the investment game, might be more appropriate for studying trusting behaviours over time.

The game data collected in this experiment reflects attitudes towards only two voices however, and we cannot be certain that these attitudes were shaped by the speakers' accents or by other characteristics of the individuals' speech. For instance, given existing literature on the topic, it would be reasonable to think that some degree of the trust attributions might be due to individual differences in voice quality and prosody.

As discussed in Section 2.2.3, research studying specifically the effect of prosody on trust attributions are limited and contradictory. The investment game methodology could help disentangle the contradictions raised by previous work, by a) studying implicit trustworthiness judgments to voice exhibiting different idiolectal characteristics, and b) systematically varying the voice and prosodic characteristics to isolate individual vocal features that contribute to these judgments. The next Chapter presents an experiment which goes a first step in this direction. The sample of accents examined was enlarged, and more than one speaker of the same accent was recorded, in order to start examining prosodic cues that might influence implicit trustworthiness judgments.

5 Experiment 2 — Accents and prosody

5.1 Introduction

The previous chapter presented a first attempt at collecting implicit measures of accent– and voice-mediated trust. Results from the investment game show that the voice of the virtual players influenced participants' investment decisions. In particular, the male Standard Southern British English (SSBE) speaker elicited particularly positive reactions when it was paired with a *generous* behaviour, and particularly negative reactions when it was paired with a *mean* behaviour. It is clear that these differential patterns of investments to the two virtual players must have been due to their voice, since this was the only feature that distinguished them. However, it is not possible to say with certainty that it was their accent that elicited a differential response. The two speakers whose voices were used in the previous experiment differed in their accent (SSBE and Liverpool), but also in other suprasegmental characteristics such as pitch, intonation, speaking rate, etc., which might all have influenced participants' reactions.

In fact, previous studies have linked vocal characteristics other than accents to personality attributions as well. For example, in a version of the investment game, where participants did not actually play the game but only indicated which possible game partner they would hypothetically trust more, O'Connor and Barclay (2017) found that they consistently selected the one with the higher-pitched voice, for both female and male voices. Conversely, a few studies have found that participants raise their vocal pitch when lying (Anolli & Ciceri, 1997; Streeter et al., 1977; Villar et al., 2013), while Zuckerman et al. (1979) found that participants perceived a lower pitch in deceptive messages, and Kirchhübel and Howard (2013) failed to find any acoustic differences — including in f0 — in the production of deceptive and truthful messages. Hughes et al. (2008) instead found a positive correlation between pitch and perceived honesty of male speakers. From a slightly different perspective, Cheang and Pell (2008) and Rao (2013) found that actors generally had a lower f0 and slower speech rate when acting sarcastic voices than when they were acting sincere voices. Similarly, sincerity in a synthetic voice was associated with greater pitch range and faster articulation rate (Trouvain et al., 2006). Slow speech rate, by contrast, was associated with less competence (B. L. Smith et al., 1975). Higher f0 was also associated with higher agreeableness (Imhof, 2010), one of the "Big Five" personality traits (McCrae, 2009; Pervin, 2001; Trouvain et al., 2006), which is connected to trust (Ashton & Lee, 2005). High f0 also elicited higher explicit cooperativeness ratings (Knowles & Little, 2016).

Some scholars found that a fast speaking rate is a feature of charismatic (Jiang & Pell, 2017), confident (Rosenberg & Hirschberg, 2005) and persuasive (Chaiken, 1979) speakers; on the contrary, Niebuhr et al. (2016) argued that fast speaking rate hinders charismatic speech, possibly because phenomena such as vowel reduction and deletion become more likely. N. Miller et al. (1976) found that faster speech rate increased persuasion, and they cite an unpublished research where it was found that faster speakers were rated as more trustworthy. Fast speech rate also elicited judgments of competence (Street, 1984; Street & Brady, 1982) and credibility (Buller et al., 1992), although this effect decreased for very fast rates of speech (Buller et al., 1992). Also, several studies on various languages found that deceptive speech was accompanied by a significantly slow speech rate (Spence, Villar & Arciuli, 2012; Rockwell, Buller & Burgoon, 1997; Vrij, Edward, Roberts & Bull, 2000; Fiedler & Walka, 1993).

Finally, voice quality — variations of which distinguish, for example, breathy, creaky and modal phonation (Klatt & Klatt, 1990) — might convey information about personality characteristics as well. For example, Laver (1968) suggested that harsh voices might be correlated with more aggressive and dominant characteristics, while breathy voices might be related to more submissive personalities. Furthermore, in another study, both hypernasal and breathy female voices — which can be symptoms of voice disorders — were rated more negatively than modal voices (Blood et al., 1979). Also, a breathy phonation mode in female voices is related to perceived femininity and attractiveness (Gussenhoven, 2016). Breathy male voices were also rated positively by female speakers (Xu et al., 2013), and the authors suggest that breathiness in male voices can signal a reduction in aggressiveness.

Talking specifically about trust attributions, Elkins and Derrick (2013) also found that a high pitch in participants' voices correlated with low trust attributions to Embodied Conversational Agents. These pitch effects also have similarities with that of a "smiling voice", as smiling tends to shorten the vocal tract and raise f0, as well as formants, which can also increase trust (Elkins & Derrick, 2013). Contrary evidence however comes from a study by Apple et al. (1979), whose questionnaire study showed that speakers with a high f0 were rated as "less truthful" (Apple et al., 1979). This study also found that speakers with slow speech rate were rated as "less truthful" and "more passive". Additionally, in a mock election scenario, participants typically voted for male candidates with lower-pitched voices (Tigue et al., 2012), and it was found that low pitch predicted actual election results (Banai et al., 2017). On the other hand, female candidates with high-pitched voices were more successful than those with low pitch (Klofstad, 2015). It should also be noted that although gender has a strong effect on pitch, with females typically having a smaller larynx and thus a higher pitch than males, there are no consistent differences in trust judgments towards men or women (e.g. Nass & Brave, 2005; Chaudhuri & Gangadharan, 2007; Boenin & Serra, 2009; Slonim & Guillen, 2010).

All these — albeit highly inconsistent — results suggest that vocal characteristics other than accent might influence trait attributions, including trust. However, no study so far has investigated the effect of vocal characteristics on implicit trust attributions. An implicit methodology aimed at eliciting actual trusting behaviours, however, might help untangle some of the confounds in previous studies. This chapter presents results of an experiment aimed at separating the effect of accent from that of other vocal cues on participants' investments in a trust game. This was achieved by using voices taken from several speakers of the same accents. In particular, the accents were SSBE, Plymouth, London and Birmingham, and these, as well as individual prosodic cues, were used in a regression model to predict monetary investments in the game. A summary of the manipulations in this experiment is shown in Table 3.

Accent	Behaviour	Participant	Speaker
SSBE	Generous	1	SSBE 3
Plymouth	Mean	1	Plymouth 3
London	Generous	1	London 3
Birmingham	Mean	1	Birmingham 3
Plymouth	Generous	2	Plymouth 1
Birmingham	Mean	2	Birmingham 1
London	Mean	2	London 1
SSBE	Generous	2	SSBE 1

Table 3: Example of the experimental conditions of two participants. Participants played 4 games in total, one with each accent, and they played 2 *generous* games and 2 *mean* games, in counterbalanced order.

5.2 Method

5.2.1 Participants

Participants were 84 native British English speakers (52 females, 32 males) aged 18–67 (median = 21, SD = 11). They were university undergraduate students who received course credit for participation or members of the public who received appropriate monetary compensation. Self-reports on participants' geographic origins were: southwest England (n = 44), southeast England (n = 20), Midlands (n = 7), Wales (n = 5), northwest England (n = 3), East Anglia (n = 2) and northeast England (n = 1). One participant was eliminated from the data-set due to being later discovered to have a non-UK English language background. The questionnaire data for one participant was not recorded due to a technical error, so that the final participant sample was comprised of 83 participants in the investment game, and 82 in the questionnaire.

5.2.2 Stimuli

Recordings were obtained of three female speakers for each of the four target accents (SSBE, Plymouth English, London English and Birmingham English), which were chosen as representatives of accents that, in literature on accent attitudes, have been associated with high trust attributions (SSBE and West Country) and low trust attributions (London and Birmingham) (Dixon et al., 2002; Fuertes et al., 2012; Hiraga, 2005). One of the main research questions for Experiment 2 was whether there would be any difference in the relative monetary investments to the regional accent and/or other vocal characteristics of the speaker, and whether this would be modulated by their behaviour within the game, i.e. with a *generous* or *mean* return of player investments. Other speakers were recorded for the questionnaire component of the experiment, with one female native English speaker from Belfast (Northern Ireland), and five female speakers of English as a second language (L2), whose native languages were Austrian German (Linz), French (Paris), Italian (Naples), Greek (Cyprus) and Mandarin (Taipei). Each speaker read 4 different blocks of 20 sentences. Two of these blocks were the same used in Experiment 1, and two more were recorded in order to obtain different sentences for all the 4 games the participants would play. The full set of sentences can be found in Appendix A. All sentences were approximately the same length (average number of syllables per sentence 16.95, SD 1.08). The recorded utterances were amplitude-normalized, and a noise-removal filter was applied.

5.2.3 Procedure

Participants sat in a computer booth, wore over-ear headphones and carried out the whole experiment on a computer. The experiment followed the same procedure as Experiment 1. Each participant engaged in four games, one for each accent condition and two for each virtual agent's behaviour condition. Thus, the accent condition was completely within-subject, since each participant played with the 4 different accents. The behaviour condition, instead, was constrained so that participants would play two *generous* games and *two* mean games. The order of appearance of the behaviour condition was counterbalanced between subjects.

The order of the blocks of sentences was randomized, and each speaker of the 4 target accents was semi-counterbalanced to appear the same number of times in the two behaviour conditions. Each game consisted of 20 rounds, which followed the same procedure as Experiment 1, with participants starting with a virtual sum of £8, hearing a sentence from the virtual game partner, deciding whether to invest any of it with it, and with the virtual player's return being displayed on the screen at the end of the round. Again, the *generous* virtual player was programmed to return 120% to 240% of the invested amount, while the *mean* virtual player returned between 0% and 120% of the invested amount. As previously mentioned (Section 4.2.3), the exact return pattern was fixed: the *generous* virtual players always returned 150% in the first round, 150% again in the second round, 180% in the third round, and so on, as explained in detail in Section 3.3.

After completing the first game, participants answered four questions about the honesty and sincerity of the virtual player they had just played with, on a 7-item Likert scale (1 = strongly disagree, 7 = strongly agree). The questions were taken from Rau et al. (2009): "This person was sincere"; "This person was interested in talking with me"; "This person wanted me to trust her"; "This person was honest in communicating with me". Again, these questions were intended to act as a safety check that participants had explicitly understood the behaviour of the virtual player and, consequently, how the game worked. The same questions were asked again at the end of the second, third and fourth game. After completing the last game and after answering these questions, participants were played one utterance from each of the speakers they had heard in the game, and asked to rate how much they liked the voice they had just heard, on a 7-item Likert scale (1 = strongly dislike, 7 = strongly like). They were then asked to state the reason behind their answer. Then they were asked to identify the accent of the speakers they heard. In contrast with Experiment 1, they were given a list of possible accents to choose from, including the four correct answers, 5 incorrect answers, and "Other" and "Don't know". The option for the SSBE accent was described as "Standard English". After this set of questions, participants were played 2 utterances from all the speakers they did not hear in the game, including the L2 speakers and the Northern Irish speaker, in random order. So, for example, if during the game they heard speakers

"SSBE 1", "Plymouth 1", "London 1" and "Birmingham 1", in this part of the questionnaire they heard speakers "SSBE 2", "SSBE 3", "Plymouth 2", "Plymouth 3", "London 2", "London 3", "Birmingham 2", "Birmingham 3", plus all the others, for a total of 14 speakers. They were asked to rate how trustworthy each speaker sounded, on a 7-item Likert scale (1 = very untrustworthy, 7 = very trustworthy). Since Experiment 1 showed that the virtual player's behaviour in the game influences subsequent explicit trustworthiness judgments of the same player, here the focus was on whether a player's behaviour influenced subsequent judgments of other speakers with the same accent, who had nothing to do with the investment game. Finally, participants completed a short background questionnaire, where they were asked their age, gender, city of origin, and what accent they spoke. This data was kept anonymous, and was used to group participants into the 7 regions described above. As a final question, participants were also asked what accent they would like a robot to have; this is discussed in full later, in Section 8.5. The total duration of the experiment was approximately 25 minutes.

5.2.4 Prosodic measures

Segmentation and labelling of the individual sound files was done with the forced alignment tool provided by the MAUS General Web service (Schiel, 1999). The transcriptions thus obtained were then used to extract prosodic measures in Praat (Boersma & Weenink, 2017) and MATLAB; specifically, the measures were mean f0, f0 range, voice quality and articulation rate. Mean f0 was calculated as the mean f0 value for each vowel, then averaged across individual utterances. Pitch range, in order to eliminate potential outliers, was calculated as the difference between the 10th and the 90th percentiles of the mean f0 value for each vowel, as in Patterson and Ladd (1999), then averaged across individual utterances. Articulation rate was calculated as syllables/second, not including pauses or other hesitations in speech (Jacewicz et al., 2009; De Jong & Wempe, 2009). Finally, H1-H2 — the difference between the first and second harmonic — was used as a measure of voice quality, as in Johnson (2002) and Garellek, Keating, Esposito and Kreiman (2013). This was calculated using

VoiceSauce (Shue, Keating, Vicenik & Yu, 2011). A summary of these prosodic measures is given in Table 4.

	Pitch (Hz)	Pitch Range (Hz)	Articulation Rate (syll/s)	H1-H2 (dB)
SSBE 1	216 (10)	29 (8)	4.11 (0.55)	10.45 (1.41)
SSBE 2	223 (10)	32 (8)	4.19 (0.52)	8.79 (1.02)
SSBE 3	224 (14)	40 (12)	4.17 (0.49)	8.43 (1.34)
SSBE average	221 (12)	34 (11)	4.16 (0.52)	9.23 (1.55)
Plymouth 1	200 (12)	30 (9)	3.81 (0.48)	9.9 (1.22)
Plymouth 2	244 (9)	28 (8)	3.67 (0.4)	7.71 (0.94)
Plymouth 3	211 (9)	27 (7)	4.21 (0.53)	7.86 (0.77)
Plymouth average	218 (21)	28 (8)	3.9 (0.52)	8.5 (1.41)
London 1	187 (8)	24 (5)	4.03 (0.54)	9.4 (0.7)
London 2	224 (9)	43 (10)	4.02 (0.45)	7.4 (0.79)
London 3	198 (8)	29 (8)	4.09 (0.52)	9.31 (1.05)
London average	203 (18)	32 (11)	4.04 (0.5)	8.72 (1.26)
Birmingham 1	165 (7)	22 (8)	3.22 (0.45)	6.19 (1.28)
Birmingham 2	241 (9)	30 (8)	4.33 (0.58)	9.15 (1.26)
Birmingham 3	209 (10)	33 (10)	4.0 (0.47)	5.99 (0.99)
Birmingham average	205 (33)	28 (10)	3.84 (0.68)	7.09 (1.86)

Table 4: Summary of prosodic measures divided by speaker (standard deviation in parentheses).

5.3 Results

5.3.1 Investment game

To determine the effects of game behaviour and vocal characteristics on investments, a mixed-effects linear model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC (Aho et al., 2014). Investment was the dependent variable; behaviour, accent, game turn, pitch range, f0, articulation rate and H1-H2 were predictors; subject and sentence block were random factors.

Effect of game behaviour

There was a main effect of behaviour, with an average investment of £6.14 to the *generous* virtual player and of £2.40 to the *mean* virtual player ($\chi^2(1) = 4290.4$, $p < .001$). There was also a main effect of game turn ($\chi^2(1) = 92.94$, $p < .001$), with higher overall investments in the second half of the game, and a significant interaction between behaviour and game turn ($\chi^2(1) = 206.82$, $p < .001$): as previously noted, investments increase in the *generous* condition in an almost linear fashion, and decrease and then increase in the *mean* condition in a u-shape fashion as the game progresses (Figure 8).

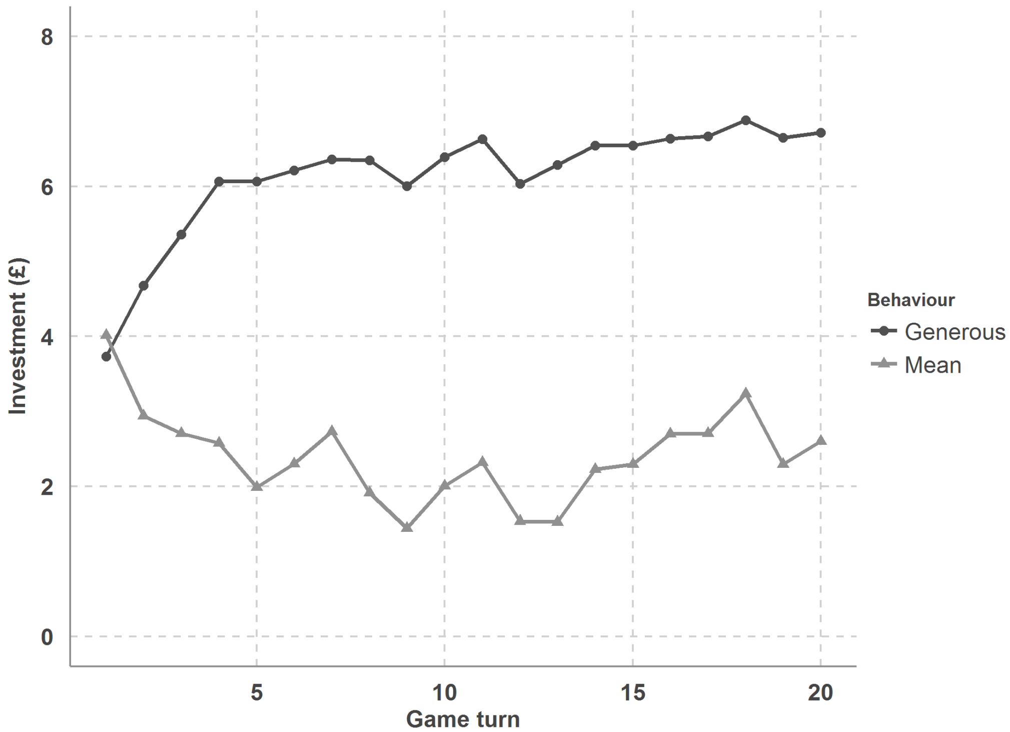


Figure 8: Average investments in the *generous* and *mean* condition.

Effect of accent

There was also a main effect of accent ($\chi^2(3) = 31.64$, $p < .001$). Post-hoc comparisons using the Tukey HSD test show that average investments to SSBE speakers were higher than the investments to Plymouth ($p = .002$), London ($p = .002$) and Birmingham ($p < .001$) speakers, with no other pairs showing significant differences. There was also an interaction between accent and behaviour ($\chi^2(3) = 14.08$, $p = .003$), but no interaction with game turn. Figure 9 shows the average investments according to behaviour and accent. Post-hoc comparisons using the Tukey HSD test showed that the investments in the *generous* condition were significantly higher with SSBE speakers than with speakers from Plymouth ($p = .024$) and London ($p = .013$), and that investments to Birmingham speakers were significantly higher than investments to London ($p = .016$) and Plymouth ($p = .029$). By contrast, in the *mean* condition, investments to SSBE, Plymouth and London speakers were all higher than the investments to Birmingham speakers ($p < .001$, $p = .022$, $p = .011$ respectively).

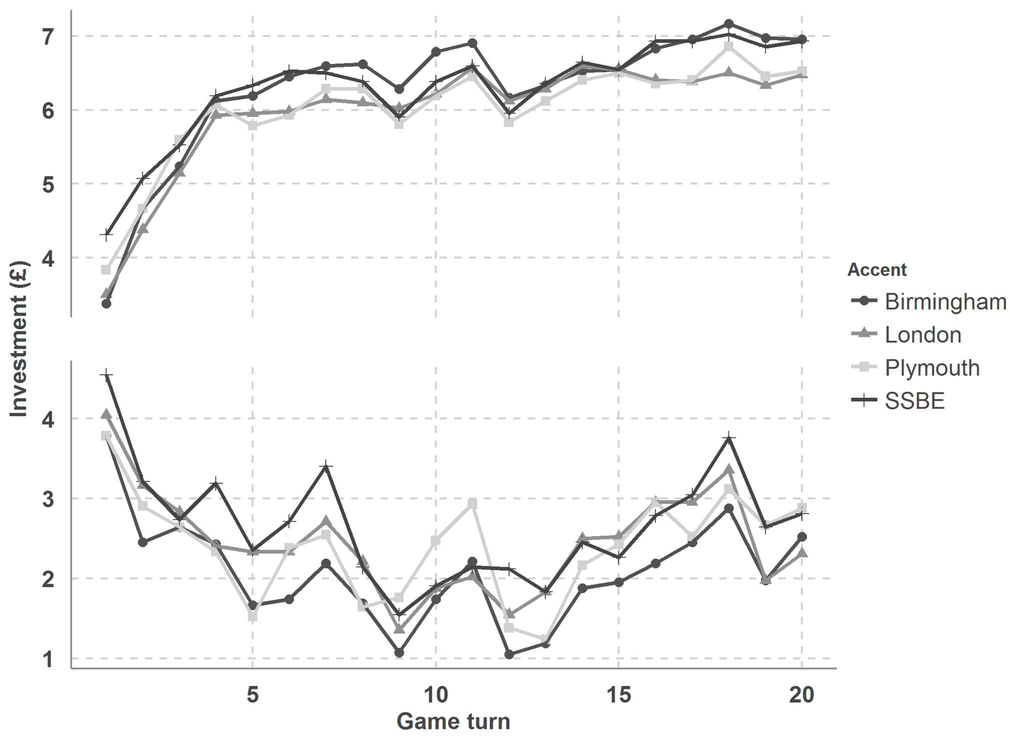


Figure 9: Average investments to the four target accents in the *generous* (top panel) and *mean* (bottom panel) conditions.

Effect of prosody

There was a main effect of f0 ($\chi^2(1) = 4.48$, $p = .034$), and the effect of articulation rate approached significance ($\chi^2(1) = 3.34$, $p = .067$, significant in terms of AIC): overall, higher pitch and faster articulation rate were associated with higher investments. These results are relatively consistent with some previous findings (e.g. Apple et al., 1979; Trouvain et al., 2006; Imhof, 2010). There were no effects of pitch range or H1-H2 on investments. There was also an interaction between f0 and game turn ($\chi^2(1) = 12.77$, $p < .001$), with higher f0 more strongly associated with higher investment earlier in the game. There was an interaction between f0 and articulation rate ($\chi^2(1) = 8.19$, $p = .017$), with greater investment for high f0 at fast rate. Finally, there was also an interaction between articulation rate and behaviour ($\chi^2(1) = 4.05$, $p = .042$), with a positive relationship between articulation rate and investment limited to the *mean* condition.

The result on articulation rate suggests that a faster rate might elicit higher trustworthiness attributions. Upon closer inspection, however, it turns out that one of the Birmingham speakers had by far the slowest articulation rate (Figure 10), as well as the lowest average f0 (Figure 11). Looking at the distribution of the articulation rates and f0s of all speakers, “Birmingham 1” is the only one to fall more than 1 standard deviation below the mean of articulation rate, and more than 1.5 standard deviations below the mean of f0. Furthermore, a 2-tailed t-test between the articulation rate of speaker “Birmingham 1” and the articulation rate of the second slowest speaker, “Plymouth 2”, was significant ($t(1092) = -17.58$, $p < .001$). Similarly, a 2-tailed t-test between the f0 of speaker “Birmingham 1” and the f0 of the second lowest speaker, “London 1”, was also significant ($t(1117.8) = -49.39$, $p < .001$). None of the other speakers had a particularly fast or high-pitched voice. This seems to suggest that speaker “Birmingham 1” was a particularly slow and low-pitched speaker in the dataset, which could be treated as an outlier in this particular set of speakers.

For this reason, a mixed-effects linear model was re-fitted excluding that particular speaker. By doing this, the effect of pitch remained the same, although its significance diminished ($\chi^2(1) = 3.23$, $p = .072$), and the effect of rate changed direction ($\chi^2(1) = 4.56$, $p = .033$).

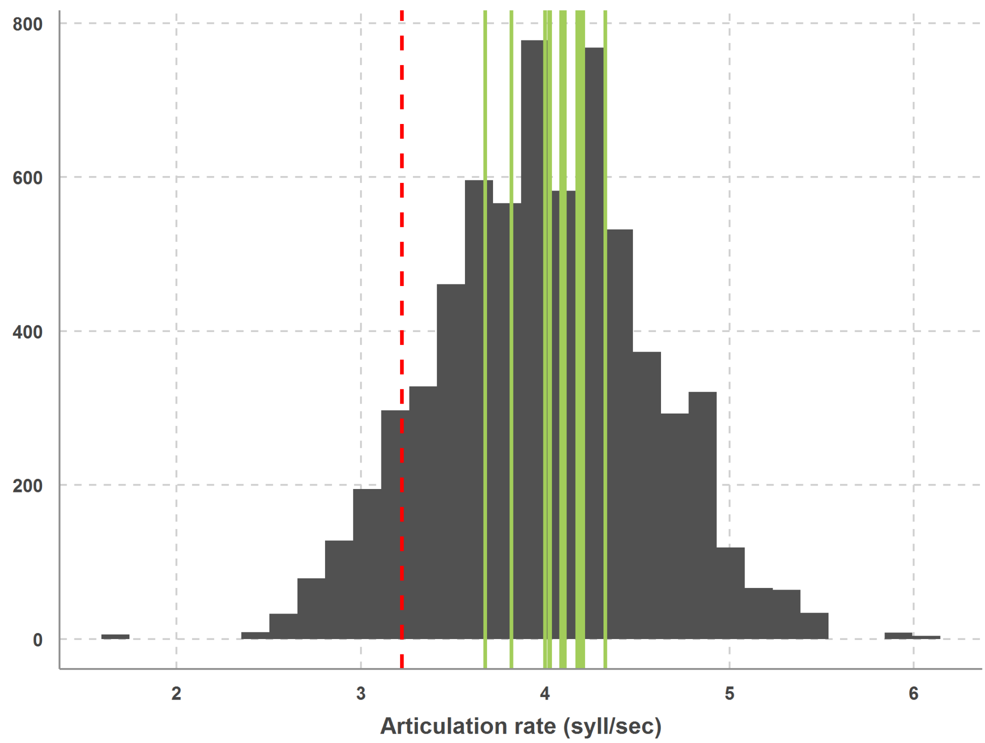


Figure 10: Distribution of articulation rates across all the utterances of the speakers used in the game. The red, dashed vertical line is the mean rate of speaker “Birmingham 1”, while the other vertical lines are the mean rates of all the other speakers.

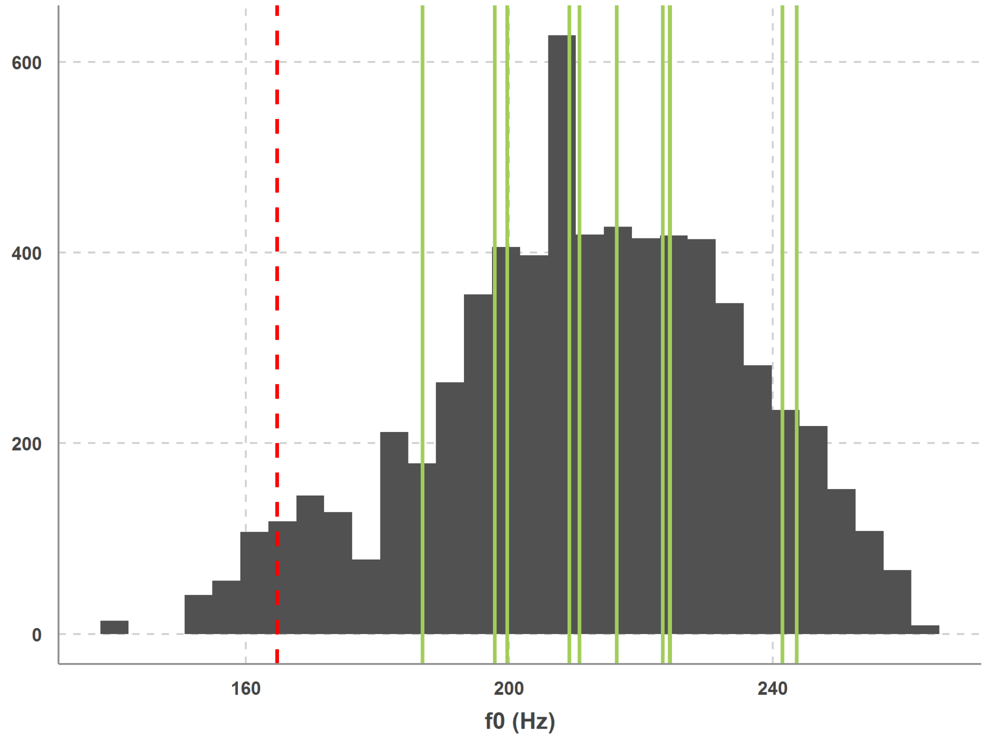


Figure 11: Distribution of f0s across across all the utterances of the speakers used in the game. The red, dashed vertical line is the mean f0 of speaker "Birmingham 1", while the other vertical lines are the mean f0 of all the other speakers.

That is to say, by excluding the extreme value, the remaining 11 speakers showed that *slower* rate predicted higher investments in the game. Also, the interaction between behaviour and articulation rate was no longer significant. The interaction between pitch and game turn was maintained ($\chi^2(1) = 5.76$, $p = .016$), as well as the main effect of accent ($\chi^2(3) = 35.20$, $p < .001$).

Cruttenden (1997) estimates the average tempo for naturally-occurring English speech to be around 6 syllables/second. Read speech generally occurs at a slower tempo than natural speech however (Hirose & Kawanami, 2002; Batliner, Kompe, Kießling, Nöth & Niemann, 1995; Laan, 1997; Dellwo, Leemann & Kolly, 2015), and Jacewicz, Fox and Wei (2010) previously found that natural speech in different varieties of American English was about 1.55 times faster than the corresponding read speech. By applying this proportion to the figure mentioned by Cruttenden (1997), the average tempo of read British English speech should be around 3.9 syllables/second. And in fact, the average articulation rate of all the utterances recorded from all the speakers in this experiment is 3.99 syllables/second. As can be seen in Figure 12, speaker "Birmingham 1" is the only one whose articulation rate mostly falls below this average. All this evidence contributes to suggesting that speaker "Birmingham 1" might have been perceived as particularly slow, at least in comparison with the other speakers that participants heard during the experiment. Thus, it seems that extreme articulation rate values might result in negative reactions. This conclusion is supported by B. L. Smith et al. (1975), who argued that extremes in both directions of manipulated articulation rate elicited negative judgments, and by Buller et al. (1992), who found that judgments of credibility decreased for extremely fast rates. Since none of the speakers in the current experiment had a particularly fast articulation rate (Figure 10), it remains to be seen whether particularly fast natural rates would elicit negative reactions in the investment game as well.

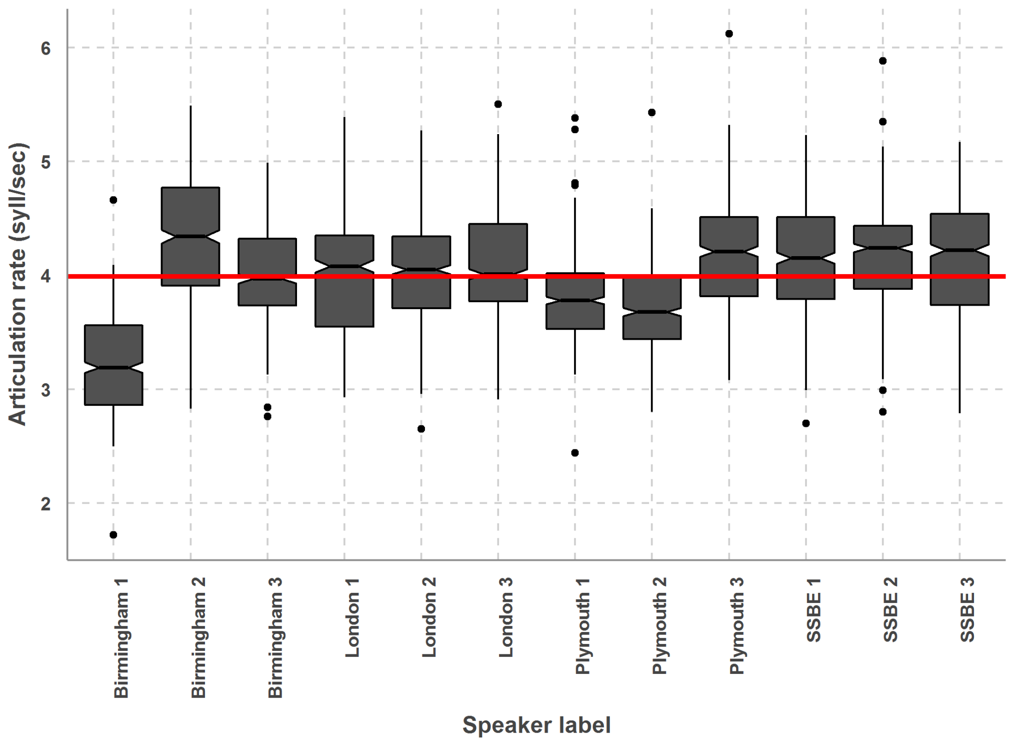


Figure 12: Median articulation rates of the 12 speakers used. Speaker ‘b1’ (Birmingham 1) has the slowest rate. The red line is the average articulation rate of all the speakers.

5.3.2 Questionnaire

Explicit honesty and sincerity evaluations

Participants were asked to evaluate the speakers they heard in the game for perceived honesty and sincerity (Rau et al., 2009). The numerical answers to these four questions were the dependent measure in a mixed-effects model: the ratings were the dependent variable, behaviour and speaker were predictors, and participants and sentence block were random effects. There was a main effect of behaviour ($\chi^2(1) = 533.88, p < .001$): as expected, participants gave higher ratings of sincerity and honesty to those players which had a *generous* behaviour in the game than those which had a *mean* behaviour (average ratings = 5.37 and 2.91, respectively). This result, together with the fact that participants invested more money with the *generous* virtual players than the *mean* players in the game, confirms that they understood how the virtual player was behaving, and reacted accordingly. There was also a main effect of accent ($\chi^2(3) = 11.17, p = .0113$). A post-hoc Tukey HSD test revealed that SSBE speakers were rated significantly higher than Birmingham speakers (average rating = 4.39 and 3.89, respectively; $p = .009$), with no other pair showing significant differences. There was also no interaction between accent and behaviour.

Voice Liking

Participants were also asked to rate how much they liked the voices they heard in the game. As expected, the speaker's associated behaviour in the game influenced this voice liking rating. A one-way ANOVA with behaviour and accent as factors revealed that both these factors influenced the liking ratings ($F(1, 324) = 13.45, p < .001$ and $F(3, 324) = 12.54, p < .001$, respectively). Participants reported liking a voice more if it had been associated with a *generous* behaviour (average = 4.35) than a *mean* behaviour (average = 3.76). Regarding the effect of accent, post-hoc comparisons using the Tukey HSD test showed that participants reported liking SSBE, Plymouth and London accents more than a Birmingham accent ($p < .001, p = .0039, p = .014$, respectively) and an SSBE accent more than a London accent ($p = .023$). There was no interaction between accent and behaviour. A summary of the liking ratings, divided by speaker, is given in Figure 13.

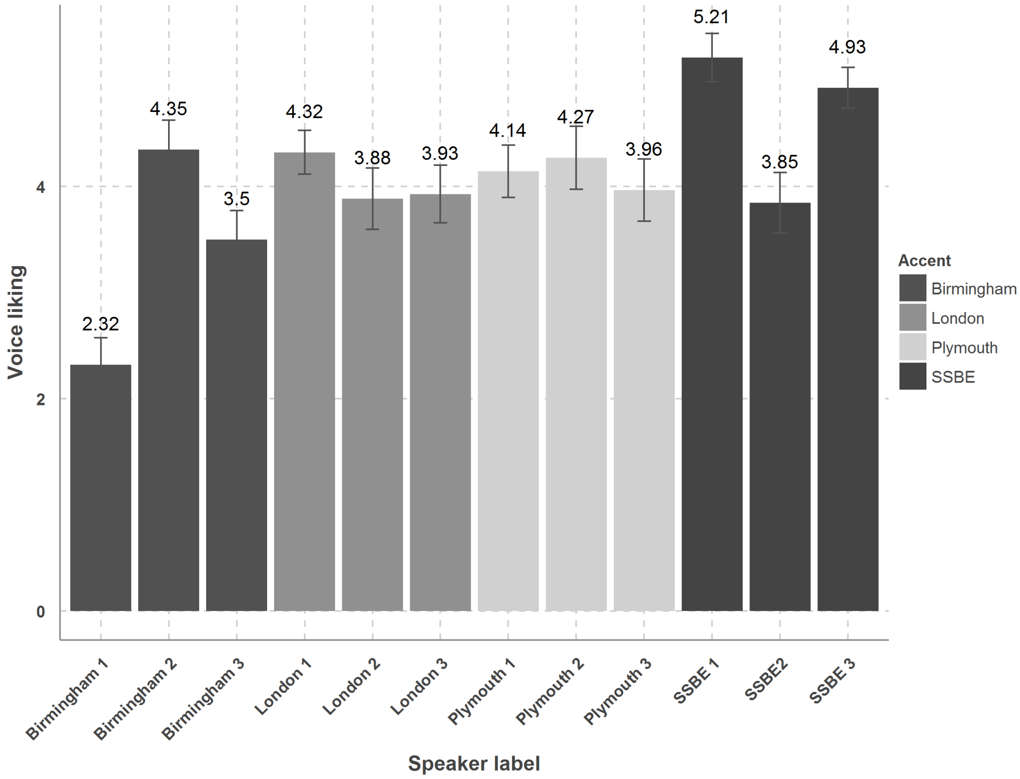


Figure 13: Mean voice liking rating of all speakers heard in the game.

Trustworthiness rating

Participants rated 14 speakers, which had nothing to do with the investment game, for trustworthiness. These included the 8 SSBE, Plymouth, London and Birmingham speakers they had not played with during the games. Figure 14 shows the mean trustworthiness rating to all the speakers recorded for Experiment 2. A one-way ANOVA analysis with the factor of accent was found to be significant ($F(3, 659) = 28.45$, $p < .001$). Post-hoc comparisons using the Tukey HSD test show that the London, Plymouth and SSBE accents were rated significantly more trustworthy than Birmingham (all $p < .001$), and that SSBE was rated significantly more trustworthy than London ($p < .001$) and Plymouth ($p = .031$). This shows a similar pattern to the investments to the four accents in the game in the *mean* condition, where Birmingham received lower investments than any other accent. However, it should be noted that participants did not rate the same speakers that they had heard in the game, but two other speakers with the same regional accent. In general, these explicit trust ratings are in agreement with prior literature, with prestigious accents such as SSBE at the higher end of the trust scale, urban accents such as Birmingham (but not London — see Figure 14 below) at the lower end, and less urban accents such as Plymouth somewhat in the middle (A. M. Wilkinson, 1965; Fuertes et al., 2012; Dixon et al., 2002; Hiraga, 2005; Crystal & Crystal, 2014). Furthermore, speakers of English as a second language received lower ratings than native English speakers (average trustworthiness rating of L1 speakers = 4.38, of L2 speakers = 4.18; $t(937.8) = 2.32$, $p = .021$), supporting previous findings (Fuertes et al., 2012; Lev-Ari & Keysar, 2010; Frumkin, 2007).

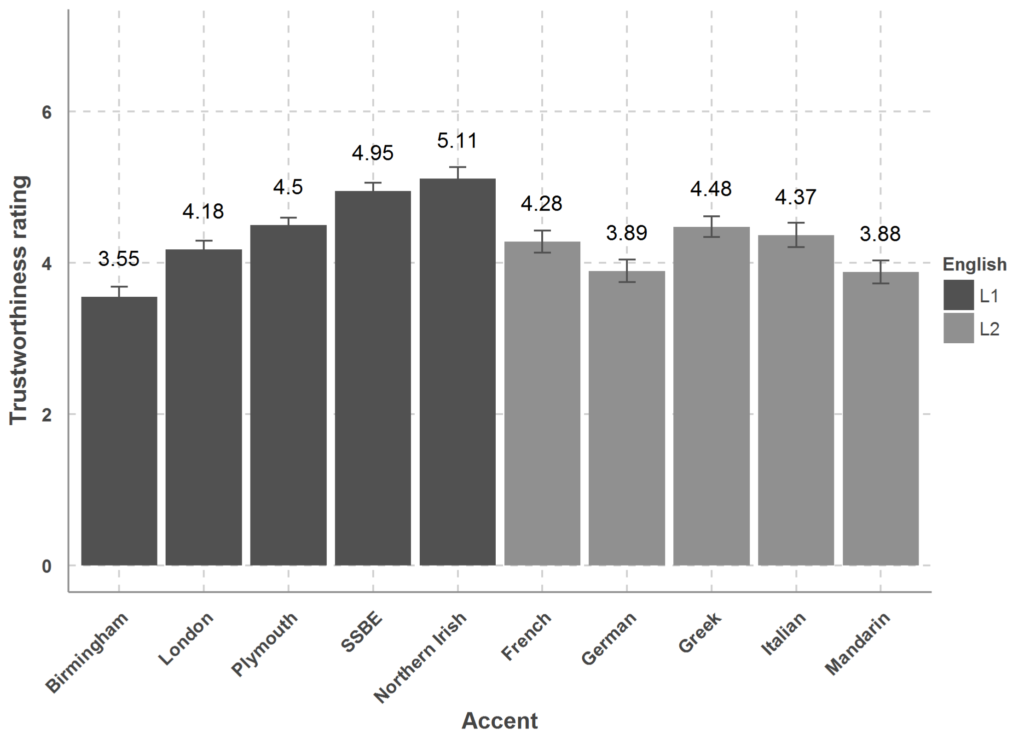


Figure 14: Mean trustworthiness rating of 10 English accents.

Contrary to the results of Experiment 1, the virtual players' behaviour in the game did not influence the trustworthiness ratings in the questionnaire ($F(1, 659) = 1.68$, $p = .19$); that is to say, accents associated with *generous* virtual players were not rated higher in the questionnaire, and accents associated with *mean* virtual players were not rated lower. This is likely due to the fact that the voices rated by participants in the questionnaire were not the same voices they heard during the game. Thus, it seems that participant do not extend the attitude that they developed towards an accented speaker to other speakers of the same accent, but rather that attitude is confined to the individual speaker who was the object of the original attitude.

Accent identification

Participants were also asked to identify the accent of the speakers they played with in the game. Given poor idenfitication rates in Experiment 1, here they were given a set of possible answers to choose from. Identification of SSBE, Plymouth and Birmingham accents were above chance level (by 53, 31 and 34 participants respectively). The London accent, on the other hand, was correctly identified only by one person. Again, this shows that trust attributions can be ascribed to speakers, even if their provenance cannot be accurately defined. As in Experiment 1, a post-hoc examination of the influence of the participant's own accent on investments with virtual players with the same or different accent was not significant ($\chi^2(1) = 0.04$, $p = .83$), thus ruling out any familiarity or "accent loyalty" effects that might have influenced participants' trusting behaviours (Giles, 1971).

5.4 Discussion

This study built upon the results of Experiment 1 both by expanding the number of accents, and by examining multiple speakers for each of the accents. As in the previous experiment, it was found that participants' investments were modulated by the regional accent of their virtual counterpart, even when the differently-accented virtual players exhibited exactly the same behaviour. It was also shown that the pattern of investments for particular accents was dependent upon the accented virtual player's behaviour. This interaction was also found in Experiment 1, albeit with a different pattern of results dictated by the different accents under analysis there. In the current experiment, the highest investments in the *generous* condition were to the SSBE- and Birmingham-accented virtual players. In the *mean* condition instead, the smallest investments were to the Birmingham-accented virtual players. In the previous study the Liverpool-accented virtual player received lower investments than the SSBE speaker in the *generous* condition, but higher investments in the *mean* condition. The pattern of results for SSBE in Experiment 1 was more similar to that of the Birmingham accent in the current study. That is, while the Birmingham speakers receive high investments in the *generous* condition, they receive the lowest investments in the *mean* condition. This provides continued support for the conclusion that the interaction between accent and behaviour is due to the incongruence between the stereotype for a particular accent and the speaker's actual behaviour. In this case, participants' first impressions about the Birmingham-accented speakers might have been of high trustworthiness. When this trustworthiness was confirmed, in the *generous* condition, participants reinforced their trusting behaviour towards these speakers. When this trustworthiness was disproved, in the *mean* condition, participants reacted more negatively to this perceived "betrayal". Also, the results regarding the Plymouth regional accent were also in line with previous research. For example, in Kristiansen et al. (1983), Devonian speakers were generally located halfway between RP and Cockney speakers in terms of trustworthiness, and indeed in the investment game Plymouth speakers were always located halfway between the most and least trusted speakers. Also, while the majority of participants reported being from the South West of England, where the Plymouth accent is geographically located, there were no in-group or "accent loyalty" (Giles, 1971) effects. That is to say, participants did not trust speakers who had an accent from their same region more.

Regarding the effect of individual differences in prosodic characteristics on the investment decisions, mean pitch (f0) and articulation rate had an effect on the trust attributions in the investment game, while pitch range and voice quality were not significant predictors in the regression analysis. Overall, higher pitch and slower articulation rate were associated with higher investments. This is consistent with Imhof (2010), who found that higher f0 was associated with higher agreeableness, a personality trait that manifests itself in a variety of behaviours, including trusting and trustworthiness (Digman, 1990; Goldberg, 1990). About rate, both Apple et al. (1979) and Trouvain et al. (2006) found that slow rate was associated with less truthfulness and less competence. While these studies used explicit evaluations of first impressions, the current findings expand on previous results by adding information about the development of voice-based attitudes over time. In particular, the interaction that was found between game turn and f0 suggests that individual prosodic cues might play a bigger role at the beginning of an interaction, but this might become secondary over time.

The greater overall trustworthiness of voices with higher pitch may be related to "Size/Frequency Code" theory (Ohala, 1983), based on higher f0 being generally associated with a smaller larynx and therefore smaller body size. As a consequence, we tend to associate lower f0 with dominance and aggressiveness, and higher f0 with friendliness and cooperativeness (Ohala, 1980; Hirschberg, 2002). In a type of interaction where trust is required, listeners might be more inclined to attribute trustworthiness to a speaker who is perceived as friendly rather than dominating. The current experiment provides evidence that this interpretation applies to female speakers however, on the basis of the speaker sample used; the experiment presented in the next Chapter will show that higher f0 also accords with trustworthiness in male speakers.

The greater trustworthiness of voices with a slow articulation rate might also be interpreted in terms of a biological code. In particular, the "Effort Code" postulates that careful pronunciation of speech can signal cooperativeness (Gussenhoven, 2002). Thus, speaking at a slower rate might signal that the speaker is willing to sacrifice information efficiency in order to ensure maximum understanding of the message. This lends support to Niebuhr et al. (2016)'s argument that a fast speaking rate hinders charismatic and persuasive speech. Articulation rate might also influence trust in terms of individual differences and in-group and accommodation theories (Giles & Powesland, 1997). The vast majority of participants in this experiment reported being from the South West of England, and speakers from this region are often described as having a slow rate of speech (Roach, 1998; Wells, 1982). Social attractiveness — a trait that has been linked to trustworthiness (Street, Brady & Putman, 1983) — was found to be influenced by relative speech similarity, including in terms of speaking rate (Street, 1984). However, results from the current experiment show no evidence of convergence towards the participants' own accents — in other words, participants' did not invest more money with speakers of the same accents as their own. Nonetheless, it is possible that in certain implicit contexts accommodation theory influences perception at a lower level than accents, that of individual speech cues.

Another interpretation of the effect of rate comes from regulatory-fit theory (Higgins, 2000). According to this theory, there are two ways of approaching a goal: promotion or prevention. Promotion-focus individuals are more oriented towards increasing the gains, whereas prevention-focused ones are concerned with minimising the losses. Furthermore, this theory can be applied to persuasion: the same message can be framed in a "promotive" or "preventive" way, and the message will have a good fit with people who approach goals in the corresponding way. The way the message "fits" with listeners is also not constant, it can be induced, and it can vary based on context. In a recent study focused on nonverbal cues, Cesario and Higgins (2008) delivered the same message in a promotive and preventive manner. The promotive message was delivered using cues such as broad gestures and fast speech rate, while the preventive message involved slow body movements and slow speech rate. Indeed, the videos were most effective for people belonging to the two types, respectively. Since this theory of persuasion is also context-dependent, if a context is more "preventive", the "preventive" message will be more persuasive. Assuming that the investment game is a preventive context, participants might have approached the goal of making as much money as possible in the interpretation of "losing as little money as possible", and thus "preventive" cues from the speakers, such as slow articulation rate, might have been more persuasive. In fact, in game theory it is well established that people are more concerned about not losing any money than gaining it (Tversky & Kahneman, 1992).

The characteristics of pitch and rate are also inherently related to the 4 accents of interest. More specifically, the SSBE speakers in the sample had the highest pitch, followed by Plymouth, Birmingham and London; the SSBE speakers had also the widest pitch range, followed by London, Birmingham and Plymouth; the SSBE speakers had also the fastest articulation rate, followed by London, Plymouth and Birmingham; finally, the SSBE speakers showed the biggest H1-H2 difference, followed by London, Plymouth and Birmingham (see Table 4 above). It is possible that the association that was found, for example, between higher pitch and higher investments, was actually a masking of the effect of the SSBE accent on the investments, which in this sample happened to be linked with higher pitch. While it is not possible, with the data collected in this experiment, to fully separate the effect of individual prosodic characteristics from the effect of accent, future studies could artificially manipulate the f0, f0 range, articulation rate and voice quality of speakers to check whether the investment patterns are actually dictated by the accent or by these individual characteristics. An experiment using this methodology is discussed in the next Chapter.

Regarding the interactions with voice characteristics, the interaction between f0 and game turn replicates a finding from Elkins and Derrick (2013), who found that the effect of pitch diminished over time. However, in their study they found that the low pitch of the trustor (the participant) predicted lower trust especially early in the interaction, while here it was found that the high pitch of the trustee (the virtual agent) predicts higher trust, especially early in the interaction. Given these differences, these two results are not incompatible, and agree that pitch might have a stronger effect initially in the interaction. Furthermore, the interaction between f0 and articulation rate found here contrasts with Apple et al. (1979), who did not find any interaction in their study, and concluded that the two measures might affect listeners' judgements independently of each other. The different methodology and technology used in the current study could explain these different results. On the other hand, the interaction between articulation rate and behaviour lends some support for Apple et al. (1979)'s conclusion, in the sense that different prosodic characteristics might independently elicit higher trust judgements in certain behavioural contexts only.

Finally, regarding the post-game questionnaires, they were intended as a comparison between an explicit measure of trust such as the trustworthiness questionnaire and the implicit trust measure brought by the investment game. In Experiment 1, the explicit trustworthiness ratings of the two accents were based on exemplars from the virtual players involved in the games. In that experiment, the behaviour assigned to the virtual player, being in the *generous* or *mean* condition, had an effect on the trustworthiness ratings of their accent. In the current study instead, the exemplars provided for accent ratings were not spoken by the virtual players, but by speakers unknown to the participants. In this case, the behaviour of the similarly-accented virtual players had no effect on the trust ratings of the accents. Thus, it is possible to conclude that the attitudes that participants have formed at the end of the game with respect to the virtual players are maintained until the questionnaire, but that their behaviour is not generalized to different speakers of the same accent.

5.5 Conclusion

The results presented in this Chapter, together with results from Experiment 1, established that participants' investments in a virtual player, a proxy for trust, were affected by their regional accents. From Experiments 1 and 2, it is possible to conclude that people implicitly trust a standard accent such as SSBE more than regional accents, particularly Liverpool, Birmingham, Plymouth and London, although the behaviour of the trustee interacts with these trusting decisions, as discussed in the next section. This is in agreement with previous research on accent attitudes in the UK, which typically found that a standard accent is evaluated more positively than regional accents, for example in terms of attractiveness and prestige (Bishop et al., 2005; Dixon et al., 2002; Fuertes et al., 2012; Giles, 1970; A. M. Wilkinson, 1965). Importantly, this effect appears to be constant over time. At the start of a game, the participants would have no experience of the trustworthiness of their game partner, and must rely upon stereotypes to make their initial investment judgments. However, it might be expected that the effect of accent might wane over the course of the game, as the expected stereotypical behaviours were replaced with actual experience. As this was not the case, it is possible to conclude that the influence of accent-related trust attribution is separate from experiential trust resulting from reinforcement learning.

Three voice features previously associated with trust — accent, high pitch and fast articulation rate — influenced participants' monetary investments. This is consistent with some previous studies on attributions of trustworthiness and associated characteristics (Giles, 1970; Dixon et al., 2002; Bishop et al., 2005; Fuertes et al., 2012, for accent effect; Zuckerman et al., 1979; Cheang & Pell, 2008; Rao, 2013, for pitch effect; B. L. Smith et al., 1975; Trouvain et al., 2006; Imhof, 2010; Apple et al., 1979, for rate effect), although the methodology employed here to establish an implicit measure of trust is radically different. Most of the studies on personality attributions to voices used actors or explicit listeners' evaluations. In the case of actors (Cheang & Pell, 2008; Rao, 2013), they might exaggerate their different personality portrayals, so as to be as different from each other as possible; the risk of this, though, is that their portrayals do not actually represent the reality of everyday interactions (Polzehl, 2015). In the case of listeners' evaluations (used e.g. in Elkins & Derrick, 2013; Apple et al., 1979; B. L. Smith et al., 1975; Imhof, 2010), it has already been discussed how explicit measures might not be very accurate in describing the complex dynamics of personality attributions (Section 1.3).

The effect of prosody found in this experiment suggests that certain suprasegmental features elicit higher trusting behaviours. However, these prosodic cues were not controlled for, but they were those that the speakers happened to have. Thus, it is possible that the results obtained here might have been an artefact of the random sample of speakers, as demonstrated by the outlier effect of speaker "Birmingham 1". The following Chapter 6 discusses an experiment in which the voice of two speakers — whose pitch and articulation rate falls within the average band of a healthy population — is artificially manipulated, in order to confirm that these two vocal cues indeed affect implicit attitude formation.

6 Experiment 3 — Manipulation of voice cues

6.1 Introduction

The previous chapter presented results of an experiment on implicit trust attributions to several speakers of different accents, where it was shown that voice characteristics other than accent influenced participants' monetary investments in the game. However, the result concerning voice characteristics was a post-hoc analysis on voices which were first and foremost recorded for their accent. Mean pitch, pitch range, articulation rate and the difference between the first and second harmonic were not chosen systematically beforehand. After demonstrating that people pay implicit attention to some of these characteristics, there is now evidence to support a more systematic and in-depth analysis of the effect of these vocal cues on trustworthiness attributions.

Scholars have artificially manipulated pitch and articulation rate of speakers to study a variety of phenomena. In a recent study, O'Connor and Barclay (2017) manipulated the pitch of male and female voices and tested its influence on general, economic-related and mating-related trustworthiness. They found that high pitch in female and male voices was associated with higher trustworthiness in economic contexts. Similarly, in a one-shot trust game, Montano et al. (2017) found that female participants trust male voices with artificially raised pitch more. In a series of early studies, Brown et al. (1975) found that slowing down speech rate and increasing pitch in male voices resulted in lower competence ratings. In terms of vocal attractiveness, which has been found to increase persuasion (Chaiken, 1979), it was shown that voices rated higher in attractiveness were associated with more favourable impressions of overall personality (Zuckerman & Miyake, 1993).

The topic of vocal attractiveness raises the issue of other evolutionary aspects of voice, namely those related to partner selection. Larynx size and vocal tract length are inversely proportional to perceived vocal pitch, and directly proportional to body size and testosterone levels in males (Dabbs & Mallinger, 1999; Fitch, 1997; Ohala, 1983; Xu et al., 2013; Titze, 1994; Abitbol, Abitbol & Abitbol, 1999). Furthermore, high testosterone levels are linked to physical aggressiveness and dominance (Puts, Hodges, Cárdenas & Gaulin, 2007; also B. C. Jones et al., 2010; Wolff & Puts, 2010; Collins, 2000; Aromäki, Lindman & Eriksson, 1999; Giammanco, Tabacchi, Giammanco, Di Majo & La Guardia, 2005; Ramirez, 2003). Since maintaining such features is costly in terms of energy and nourishment, it has been suggested that masculine traits such as low vocal pitch might be a "honest signal" of a male's physical health and reproductive success (Apicella, Feinberg & Marlowe, 2007; Feinberg, Jones, Little, Burt & Perrett, 2005; Puts, Gaulin & Verdolini, 2006; Folstad & Karter, 1992), and that vocal pitch contributes to male partner selection (Puts et al., 2006). Conversely, higher female vocal pitch is correlated with long-term health (Vukovic, Feinberg, DeBruine, Smith & Jones, 2010) and oestrogen levels (Abitbol et al., 1999; Pipitone & Gallup, 2008), which in turn signal a healthy reproductive system (Alonso & Rosenfield, 2002; Feinberg, Jones, DeBruine et al., 2005; Van Borsel, De Cuypere, Rubens & Destaerke, 2000; Baird et al., 1999; Lipson & Ellison, 1996). Thus, a feminine trait such as a high vocal pitch can be similarly considered a "honest signal" of female fecundity (Pipitone & Gallup, 2008; Bryant & Haselton, 2009). This contributes to explaining the numerous findings on vocal attractiveness, specifically that men prefer high-pitched female voices and that women prefer low-pitched male voices in romantic contexts (Re et al., 2012; Saxton et al., 2006; Hodges-Simeon et al., 2010; O'Connor & Barclay, 2017; Riding et al., 2006; Borkowska & Pawlowski, 2011; Feinberg et al., 2008; B. C. Jones et al., 2010).

Following this line of thought, it has been argued that men's dominance is related to their physical strength (Sell et al., 2009) and thus people might be more inclined to follow a low-pitched male leader (Tigue et al., 2012; Banai et al., 2017; O'Connor et al., 2014; Klofstad, Anderson & Peters, 2012), while women's dominance is related to social skills such as cooperativeness (McAllister, 1995; Nettle & Liddle, 2008), which have been linked to high vocal pitch (Imhof, 2010; Knowles & Little, 2016). This would explain findings showing that people find low-pitched men more dominant (Puts et al., 2007; Wolff & Puts, 2010; Collins, 2000), but it would not explain findings of perceived dominance of low-pitched female voices (B. C. Jones et al., 2010). Perhaps, as it has been suggested, the two vocal cues refer to two different aspects of dominance: specifically, *physical* dominance would be linked to a low voice pitch, and *social* dominance to a high voice pitch. If this were the case, in a game such as the investment game, where mutual cooperation leads to mutual benefit, we would expect participants to invest more money in the compatriots exhibiting a signal of cooperativeness – trust that the compatriot is going to cooperate – such as high pitch. However, these theories so far are missing two critical points in their argument. The first is that, while "honest signals" of reproductive success might still apply today in romantic contexts, humans have gone through centuries of social evolution, and modern societies (at least those where participants in recent experiments come from) are not supposed to have the same gender roles, for example in terms of dominance, that our ancestors had (Diekman & Eagly, 2000; Eagly, Wood & Diekman, 2000). Secondly, as previously mentioned in Section 1.4, we most likely do not need to activate a "fight or flight" response every time we hear a new speaker, but rather we change our judgment from first impressions to impressions informed by a speaker's behaviour.

Contrarily to vocal pitch, articulation rate is not determined by body size or hormones, but rather it is context and speaker-specific. In fact, we use rate in our speech differently based on the formality, affect, mood, communication style, etc. of the situation we are in. Articulation rate is also affected by individual speaker characteristics such as age, geographic region of origin, native language, socioeconomic status etc. (Jacewicz et al., 2009; Byrd, 1994; Hewlett & Rendall, 1998; Verhoeven, De Pauw & Kloots, 2004). It also links back to information processing theory, according to which it is better to put as much information as possible in as little speech as possible. Confirming this theory, it was famously demonstrated that frequent words (such as pronouns or prepositions) are shorter than less frequent words (Zipf, 2012; Piantadosi, Tily & Gibson, 2011). Stemming from this, one could argue that speaking faster allows to include more information in less time, and should therefore be the preferred method of communication. However, speaking fast can also give rise to phenomena such as sound reduction and elision (Gay, 1981; J. L. Miller, 1981; Uchanski, Choi, Braida, Reed & Durlach, 1996), which might impair speech comprehension. After all, speakers speak in order to be understood, so it should be in their interest to convey as clear a message as possible. From this point of view, a clear, slow articulation rate might signal a speaker's competence and knowledge and might be trusted more than a fast speaker, who might be perceived as trying to prevent counter-arguing (N. Miller et al., 1976). On the other hand, speaking faster might also be seen as a corollary of greater conversational effort (Gussenhoven, 2002), and people might be more inclined to trust someone who is putting effort and energy into the conversation. Results from Experiment 2 support the former hypothesis, but they are somewhat marginal with respect to the stronger effect of pitch. The current chapter will discuss whether specific rate manipulations influence participants' implicit trust.

From the literature mentioned before, it seems that there is enough evidence to claim that vocal characteristics affect judgments of economic trustworthiness/cooperativeness, although their origin, and whether or not they are maintained in time, is uncertain.

While it has been shown that very little time is necessary to form first impressions (Zebrowitz & Collins, 1997; McAleer et al., 2014; Willis & Todorov, 2006), it is also true that exposure time increases accuracy and confidence in personality judgements (Blackman & Funder, 1998; Ambady et al., 1999; Carney et al., 2007; Willis & Todorov, 2006). Furthermore, it is also well known that interactants will adapt to each other's behaviour, gesture, and speech over time (Accommodation Theory, see Burgoon et al., 2007; Giles & Powesland, 1997); because of this continuous adaptation, behaviours dynamically change in response to affect, speaking partner, and environment (Elkins & Derrick, 2013). Evidence of attitude change also comes from Contact Theory: intergroup interactions, real or imagined, have been found to reduce initial prejudice (Pettigrew & Tropp, 2006; Crisp & Turner, 2009; Wright et al., 1997), both in adults and in children (Cameron et al., 2006). Similar trends have been found for accents (Adank et al., 2013). Thus, while it is clear that personality attributions change with experience, it is still not clear how voice-based first impressions change when interacting with behaviour and experience. The experiments presented in this thesis can help elucidate the matter. If a speaker's voice elicited immediate implicit trusting behaviours by itself, we would see differences in investments to the different virtual players from the first round of the game, when participants have no information about the virtual player's actions. This was not the case for the first experiment, where there were no differences in investments to the SSBE and Liverpool accent during the first round of the game. In the second experiment, instead, differences in investments to the different speakers emerged from the first round. More crucially though, the judgments were built and reinforced over time, and differed based on the virtual player's actions. This chapter will discuss the effect of vocal cues on trust more in depth, and whether they have an effect on first impressions, their effect is built over time, or both.

This chapter presents results from an experiment aimed at studying more in depth the effect of speakers' prosodic characteristics – pitch and articulation rate – on implicit judgments of trustworthiness.

6.2 Method

6.2.1 Participants

Participants were 120 native British English speakers (100 females, 20 males) aged 18-46 (median = 19, SD = 3.67). They were university students who received course credit for participation. Their geographic origins were reported as: southwest England (n = 69), southeast England (n = 26), Midlands (n = 11), East Anglia (n = 5), Wales (n = 4), Yorkshire (n = 2), northwest England (n = 1), northeast England (n = 1) and the Channel Islands (n = 1).

6.2.2 Stimuli

Two of the speakers who had previously been recorded for the first and second experiment – the male SSBE speaker from Experiment 1 and one of the female SSBE speakers from Experiment 2 – were used as baseline for the voices of the current experiment. The f0 and articulation rate of each recorded utterance was then manipulated using the software for phonetic analyses Praat (Boersma & Weenink, 2017). In particular, the pitch was raised or lowered by 15% (*high* or *low* pitch) and the articulation rate was sped up or slowed down by 10% (*fast* and *slow* rate). Pitch and rate were manipulated using the Pitch-Synchronous Overlap Add (PSOLA) algorithm (Moulines & Charpentier, 1990). This method allows to compress or expand the time or frequency domains of the speech signal with very few changes in the other domain, and has been widely used in speech perception studies (e.g. Re et al., 2012). A summary of the original and manipulated vocal pitch and rate can be found in Table 5.

	Female	**Male**
Original f0 (Hz)	229	122
Raised f0 (Hz)	265	141
Lowered f0 (Hz)	198	107
Original rate (Syll/sec)	4.3	4.1
Increased rate (syll/sec)	4.6	4.38
Decreased rate (Syll/sec)	4.04	3.95

Table 5: Original and manipulated vocal properties.

6.2.3 Procedure

The experiment followed the same procedure as the previous two experiments. Each participant engaged in two games, one for each pitch, rate, speaker and behaviour condition (*generous*/*mean*), with a different block of 20 utterances heard for each version of the game. The experiment was counterbalanced in a 2 (pitch: *high* or *low*) by 2 (rate: *fast* or *slow*) by 2 (behaviour: *generous* or *mean*) within-subject design. Furthermore, the gender of the speaker was different in each game played, and the order of appearance of male or female voices was randomized. The game rounds proceeded as in the previous experiments.

After completing the 20 rounds of the game participants were asked to complete a short questionnaire. Firstly, participants were played 8 utterances taken from the full set of vocal manipulation conditions, and were asked to rate them for naturalness using a Likert-type scale (1 = very unnatural, 7 = very natural). In order to compare the perceived naturalness of the manipulated utterances with the original ones, participants were also played 2 original utterances, together with the manipulated ones. The order of presentation of all the utterances was randomized across participants. Finally, participants completed a short background questionnaire, where they were asked their age, gender, city of origin, and what accent they spoke. As a final question, participants were also asked what accent they would like a robot to have; this is discussed in full later, in Section 8.5. This data was kept anonymous, and was used to group participants into the 5 regions described above. The total duration of the experiment was approximately 20 minutes.

6.3 Results

6.3.1 Investment game

To determine the effects of game behaviour and vocal manipulations on investments, a mixed-effects linear model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC (Aho et al., 2014). Investment was the dependent variable; behaviour, game turn, f0, articulation rate and speaker gender were predictors; participant id and sentence block were random factors.

There was a main effect of behaviour, with an average investment of £5.6 to the *generous* virtual player and of £2.8 to the *mean* virtual player ($\chi^2(1) = 2007, p < .001$). There was also a main effect of game turn ($\chi^2(1) = 193.45, p < .001$), with higher overall investments in the second half of the game, and a significant interaction between behaviour and game turn ($\chi^2(1) = 73.013$, $p < .001$): as in previous experiments, investments increase in the *generous* condition and decrease and then increase in the *mean* condition as the game progresses (Figure 15).

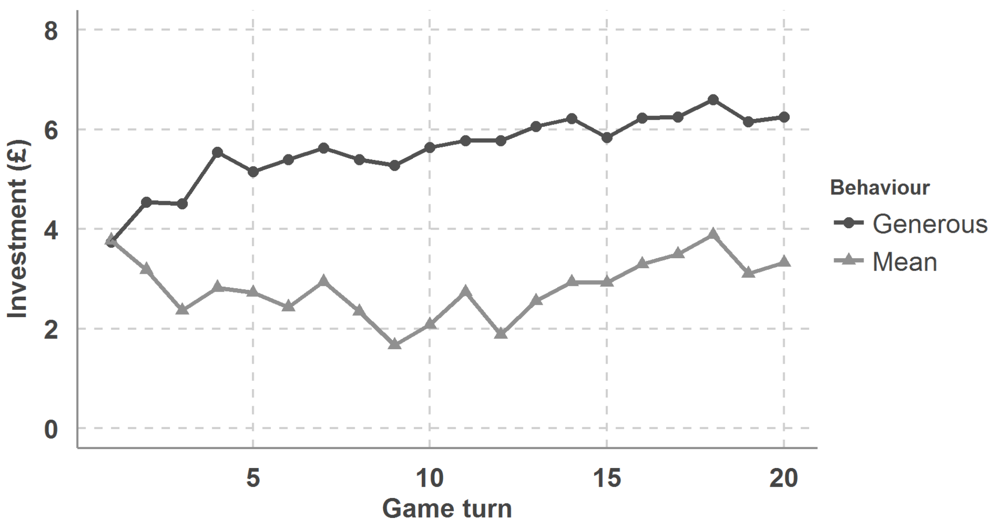


Figure 15: Average investments in the *generous* and *mean* condition.

There was also a main effect of pitch ($\chi^2(1) = 44.58$, $p < .001$), with higher overall investments with high pitch, a main effect of speaker ($\chi^2(1) = 4.63$, $p = .031$), with higher overall investments for the female speaker, and a main effect of rate ($\chi^2(1) = 5.25$, $p = .022$), with higher overall investments in the slow rate condition. There was also an interaction between game turn and articulation rate ($\chi^2(1) = 14.71$, $p < .001$); as can be seen from Figure 16, participants invest higher amounts in the slow rate condition in the second half of the game.

There was also a marginally significant three-way interaction between pitch, speaker and behaviour ($\chi^2(4) = 9.33$, $p = .053$). As shown in Figure 17, while investments in the *generous* condition remain relatively similar, in the *mean* condition the highest investments go to the female high-pitched voice, followed by the male high-pitched voice, the female low-pitched voice and finally the male low-pitched voice. A post-hoc mixed-effects model split in the two behaviour condition, and using pitch and speaker gender as predictors, confirmed this trend. Specifically, there was no main effect of pitch or speaker, and no interaction

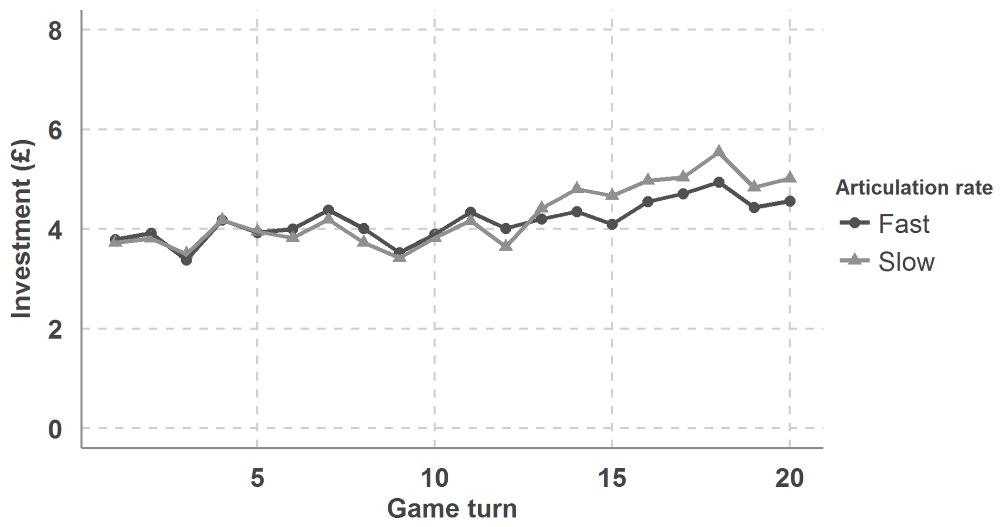


Figure 16: Investments over time in the two rate conditions

between the two, in the *generous* condition, while there was a main effect of pitch in the *mean* condition ($\chi^2(1) = 4.98$, $p = .026$).

Finally, there was a three-way interaction between pitch, rate and game turn ($\chi^2(2) = 25.7$, $p < .001$). As shown in Figure 18, investments in the slow rate conditions increase rapidly in the second half of the game, but only if paired with high pitch. This was confirmed by a post-hoc mixed-effects model fitted with game turn and rate as predictors, split in the two pitch conditions. In the high pitch condition, there was a main effect of game turn ($\chi^2(1) = 150.23$, $p < .001$) as well as an interaction between game turn and rate ($\chi^2(1) = 37.1$, $p < .001$). On the other hand, in the low pitch condition there was only a main effect of game turn ($\chi^2(1) = 88.76$, $p < .001$).

In order to determine whether speaker judgments were formed from the very beginning of the game, a post-hoc mixed-effects model was fitted only on the data regarding the first round of the game. Investment was the dependent variable; f0, articulation rate and speaker gender were predictors, and participant id and sentence block were random factors. Behaviour was not included as a predictor, since information about the virtual player's actions

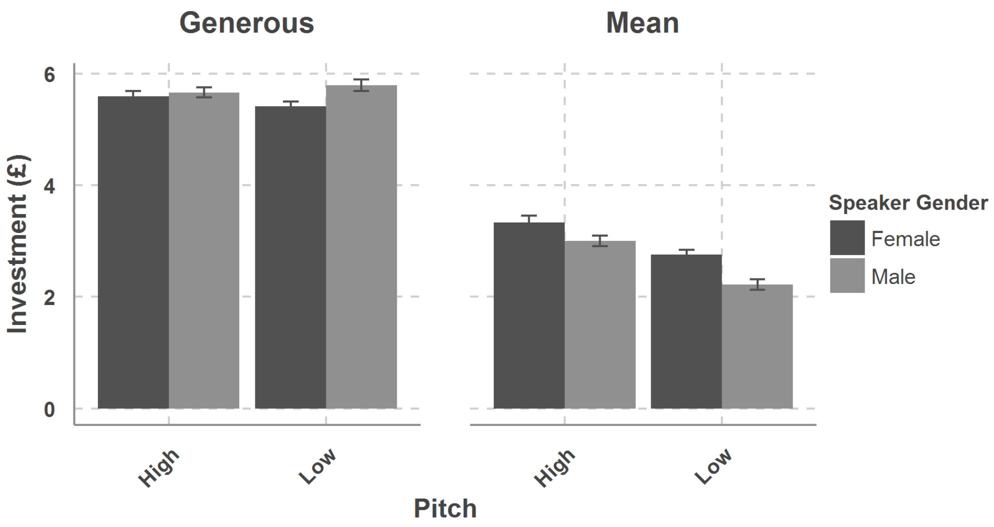


Figure 17: Investments divided by pitch and gender

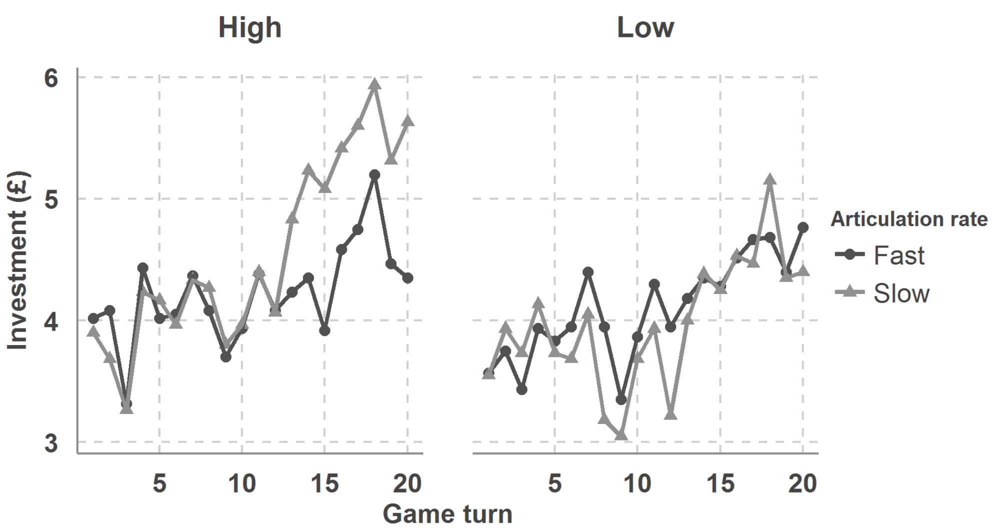


Figure 18: Investments over time divided by rate, in the high (left plot) and low pitch (right plot) conditions

is only discernible from the second round of the game on. Only a model with f0 as predictor provided significantly better fit to the data than the null model ($\chi^2(1) = 8.72$, $p = .003$). Thus, it seems that pitch was the only vocal cue that elicited higher investments from the beginning, with participants investing more in the high-pitched speakers from the first round of the game.

6.3.2 Naturalness questionnaire

Due to a technical error, the naturalness ratings of 8 participants were not recorded; therefore, the sample size for this part of the experiment is of 112 individuals.

To determine if there were any systematic differences in the naturalness ratings of the various utterances, a mixed-effects linear model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC. The naturalness rating was the dependent variable; modified pitch and rate were predictors; participant id, sentence id and speaker gender were random factors. Since this questionnaire was meant as a verification that the speech manipulations did not affect the quality of the signal, speaker gender was considered a random factor for this analysis.

Neither pitch nor rate had an effect on the naturalness ratings ($\chi^2(2) = 1.22$, $p = .54$ and $\chi^2(2) = 1.97$, $p = .37$, respectively), although there was a non-significant tendency to rate slowed-down speech as less natural (mean naturalness rating for fast rate = 4.04, mean naturalness rating for slow rate = 3.63).

6.4 Discussion

This experiment was intended as a deeper examination of the effect of prosodic cues on trust attributions, after the results from Experiment 2 indicated that pitch and rate might contribute to them. Agent's behaviour, time, higher pitch, slower articulation rate and female voice all contributed to increased trust attributions in the investment game.

6.4.1 Effect of gender

Regarding the difference between female and male voices *per se*, female voices were perceived as more trustworthy than male ones. Given the highly controversial and contradictory nature of research on trust attributions to genders, this result necessarily supports some previous studies and contrasts others (for a review, see Buchan et al., 2008). While studies on personality attributions based on voice mostly rely on explicit measures, gender studies have used implicit measures of trustworthiness similar to the investment game used here. Thus, it seems reasonable to compare the current result with previous studies employing similar economic games. In the past, several studies used one-shot investment games with partners whose gender was known to the participants. Their results either showed that females were perceived as more trustworthy than males (Buchan et al., 2008; Eckel & Wilson, 2005), or that there were no differences in trustworthiness based on the partner's gender (Chaudhuri & Gangadharan, 2007; Eckel & Wilson, 2003). Also, participants in the one shot investment game used in Boenin and Serra (2009) returned more money to partners of the same sex. As noted earlier, the majority of participants who took part in the current experiment were females, so it is possible that the higher trustworthiness that was found for the female voice was actually masking a same-sex preference. Since mixed-effects models are robust to outliers and to groups of different sizes (Quené & Van den Bergh, 2008), the final model was compared to a post-hoc model which included participant's gender as a predictor of investment. The second model did not result in a better fit for the data ($\chi^2(1) = 0.032$, $p = .86$), nor were there any interactions between participant gender, speaker gender, and pitch, thus suggesting that female and male participants behaved similarly in the game. As such, there is consistent evidence to suggest that females might be implicitly perceived as more trustworthy than males, and that these attributions might be generalisable across information channels: while in the present study the auditory channel was used, participants in previous research were either given a photograph of the game partner (Eckel & Wilson, 2005), or could imply their partner's gender by their name (Buchan et al., 2008).

As mentioned earlier, trust is a lower-level trait of agreeableness (Ashton & Lee, 2005), one of the Big Five personality traits (McCrae, 2009; Pervin, 2001). A recent research on Theory of Mind (Nettle & Liddle, 2008) argued that the higher agreeableness score associated with women in their study might be explained culturally, since females need(ed) social bonds more than males, for example when bringing up offspring. On the contrary, men's reproductive success depends on their status (Pollet & Nettle, 2008; Nettle & Liddle, 2008). Men are therefore led to increase their status, even at the cost of sacrificing their agreeableness (Nettle & Liddle, 2008). It is therefore plausible that a characteristic that is both socially (Borkowska & Pawlowski, 2011; Puts, Barndt, Welling, Dawood & Burriss, 2011; Ko, Judd & Blair, 2006) and physically (Titze, 1989) linked to agreeableness/femininity, such as high pitch, might be consistently linked to implicit trustworthiness judgments.

6.4.2 Effect of pitch

A high-pitched voice overall increased participants' trust attributions, similarly to results from Experiment 2. This is consistent with previous research using economic games (O'Connor & Barclay, 2017; Montano et al., 2017), cooperativeness ratings (Knowles & Little, 2016) and some trustworthiness ratings (O'Connor & Barclay, 2017). On the other hand, the current result contrasts with previous research on the effect of pitch on trust attributions as measured through leadership (e.g. in voting behaviour: Tigue et al., 2012; Klofstad et al., 2012) and other trustworthiness ratings (Apple et al., 1979; Tsantani, Belin, Paterson & McAleer, 2016). None of these studies measured trust as monetary investments in the game, so it is possible that we all measured different facets of the same phenomenon, as previously suggested (O'Connor & Barclay, 2017; Montano et al., 2017; Berg et al., 1995). Furthermore, economic trust and cooperativeness have been linked to cognitive forms of interpersonal trust, which refer to traits such as reliability, integrity, honesty, and fairness of a trustee (McAllister, 1995; Nettle & Liddle, 2008). On the other hand, leadership trust has been linked to status and dominance (O'Connor et al., 2014), and low vocal pitch in both male and female voices has been shown to be a sign of dominance (B. C. Jones et al., 2010; Puts et al., 2007; Wolff & Puts, 2010; Collins, 2000). Dominance is also related to evolutionary aspects such as mating behaviours; in this context, low vocal pitch is an "honest signal" (Fraccaro et al., 2013; Hodges-Simeon, Gurven, Puts & Gaulin, 2014) of bigger body size and high testosterone levels (Dabbs & Mallinger, 1999; Fitch, 1997; Ohala, 1983; Xu et al., 2013), which predict reproductive success for males (Apicella et al., 2007; Feinberg, Jones, Little et al., 2005). This additionally supports the previously mentioned hypothesis that the greater trustworthiness of voices with higher pitch may be related to the "Size/Frequency Code" theory (Ohala, 1983, see Section 5.4). This suggests that, from an evolutionary point of view, a smaller larynx could be perceived as non-threatening (Ohala, 1980; Fitch & Giedd, 1999; Hodges-Simeon et al., 2014). This mechanism is widely used in nonhuman animals (Fitch, 1999; Hauser, 1993), but evidence shows that human listeners are still able to make very accurate judgments about speakers' body sizes as well (D. R. R. Smith, Patterson, Turner, Kawahara & Irino, 2005). Thus, it is likely that the investment game is linked to social, cooperative trust, rather than leadership trust, and that high vocal pitch in this context might be implicitly interpreted as a predisposition to social cooperation. Conversely, low vocal pitch might be a correlate of leadership trustworthiness. In Experiment 2 only female voices were used; by extending those results to male voices, it is possible to conclude that higher f0 increases implicit perceived trustworthiness in both genders.

The effect of speaker gender is consistent with previous research that found that high pitch in female voices elicited higher trustworthiness judgments (O'Connor & Barclay, 2017), and with results from Experiment 2 in this thesis, where female speakers with high pitch received higher amounts of money in the investment game. Participants invested more money with female speakers who had a naturally high pitch in Experiment 2 – compared to other natural voices – and with a female speaker who had an artificially increased pitch – compared to the same voice with an artificially decreased pitch. Thus, it appears that the effect of pitch for female voices might be independent from other voice characteristics, such as voice quality or pitch range, which remained the same in the two manipulated versions of the voice used in the current experiment.

High-pitched female voices are also consistently linked to attractiveness judgments, which relate to mate selection (Borkowska & Pawlowski, 2011; Feinberg et al., 2008). For example, it was found that male listeners consider altered high-pitched female voices more attractive (Fraccaro et al., 2013), even when their pitch is manipulated to be very high (Re et al., 2012). Since it has been argued that, over evolutionary time, women have maximised their survival by strengthening social bonds (Nettle & Liddle, 2008), it seems plausible that a "non-threat signal" such as high pitch might have evolved to be a corollary of high trustworthiness as well.

This result also supports the relation between high pitch and positive traits on male voices reported in previous studies. For instance, McAleer et al. (2014) found that male voices with a relatively high pitch (close to an average female pitch) were perceived as having more valence. In a comparison of charismatic and non charismatic speakers, both Rosenberg and Hirschberg (2005) and Niebuhr et al. (2016) found that high pitch was one of the characteristics that stood out in the most charismatic male speaker's acoustic profile. Also, in some early vocal manipulation studies, it was found that raised pitch in male voices was rated more positively (Brown, Strong & Rencher, 1973; Brown, Strong & Rencher, 1974). However, a few studies reported opposite results of preferences towards male speakers with low pitch. In Apple et al. (1979) high-pitched, slow male voices were rated particularly "non truthful"; also, male and female participants voted for male candidates with low pitch voices, both in a simulated election scenario (Tigue et al., 2012), and in actual elections (Banai et al., 2017).

It is also possible that individual differences account for the effect of speaker gender and pitch. In fact, previous studies suggest that male and female participants might prefer a higher or a lower-pitched voice based on the speaker gender, with female speakers generally preferring low-pitched male voices, and male participants favouring high-pitched female voices (Re et al., 2012; Saxton et al., 2006; Hodges-Simeon et al., 2010; O'Connor & Barclay, 2017; Riding et al., 2006). The participant sample for this experiment does not allow to make inferences about such individual differences, given the disparity of female and male participants. However, if female participants overall preferred a low-pitched male voice, as suggested in the research mentioned above, we would see an interaction between pitch, speaker gender and participant gender in the present data, which we do not. Thus, it seems that vocal preferences might cease to be gender-specific in implicit tests of trustworthiness. The absence of an interaction between speaker gender and pitch also suggests that vocal signals of trust might be the same across genders. At the same time, women's f0 are generally higher than men's, so the results suggest that people might find high-pitched male voices and high-pitched female voices trustworthy. Thus, pitch might be relative to stereotype for each gender, rather than an absolute cue.

6.4.3 Effect of articulation rate

A slower articulation rate predicted higher monetary investments in the game. This supports previous findings on the characteristics of charismatic speech (Niebuhr et al., 2016), while contrasting with others (Jiang & Pell, 2017; Rosenberg & Hirschberg, 2005). The effect of rate found here also supports results from Experiment 2, where it was shown that, after removing an outlier, slower articulation rate predicted higher investments in the game. Again, results from the present experiment expand from the results of the previous one, where the speakers happened to have a naturally fast or slow rate in comparison with other speakers. Here, a speaker's rate was manipulated in order to be faster or slower while maintaining all other voice characteristics constant. Thus, it seems that also the effect of articulation rate might be independent of other voice characteristics. Interestingly, in the present experiment there was an interaction between pitch, rate and game turn, with the combination high pitch and slow rate eliciting higher investments particularly in the second half of the game. This seems to suggest that, while the two vocal cues might overall evoke positive impressions independent of each other, a combination of the two might increase trust after an initial learning period.

As previously mentioned in Section 5.4, the results on articulation rate might be explained in terms of the "Effort Code" account of phonetic universals (Gussenhoven, 2002). This suggests that careful pronunciation of speech, for example through a slow articulation rate, can signal cooperativeness. It could also be due to low-level speech similarities between participants and speakers, since the majority of participants in the present experiment reported being from the South West of England, where local accents have generally a slow rate of speech (Roach, 1998; Wells, 1982; Giles & Powesland, 1997). Another interpretation might come from regulatory-fit theory (Higgins, 2000), described in Section 5.4.

6.4.4 Effect of time

The effect of prosodic cues over time is consistent with results from previously presented experiments, suggesting that trust attributions are dependent on speaker's behaviour, and at the same time participants maintain their preference towards one or another speaker. In Experiment 2 it was found that the effect of vocal pitch on investments decreased over time, as the game progressed. Here it appears that the effect of pitch remains constant throughout the game, while the effect of rate takes some time to develop fully.

6.4.5 Naturalness ratings

Regarding the naturalness ratings, it is interesting to see how none of the stimuli, including the original utterances, were rated as particularly high on naturalness. On the contrary, the mean rating for all the stimuli was around 4, on a scale where 1 corresponded to "very unnatural" and 7 to "very natural". Since there were no differences between the natural, original utterances and the manipulated ones in their perceived naturalness, it can be excluded that the digital alterations corrupted the audio signal to the point of becoming artificial-sounding. Participants heard utterances belonging to all the stimuli conditions, and were asked to "rate how natural the speaker's voice sound[ed]". The instructions had this wording in order to prevent participants from rating the *semantic content* of the utterances, and such instructions are common in speech perception studies (to name a few: Tsantani et al., 2016; Fraccaro et al., 2013; Babel et al., 2014). It would be possible to further investigate this doubt by masking the semantic content of these utterances, for example by reversing them (Munro, Derwing & Burgess, 2010; Scherer, Ladd & Silverman, 1984), and ask participants to rate these non-utterances in terms of naturalness. However, given that the main purpose of the naturalness questionnaire was to make sure that the manipulations did not produce artificial or non-human sounding stimuli, resulting in the potential loss of ecological validity, and that the instruction wording is a common used methodology, I am satisfied with the current results.

6.5 Conclusion

This chapter presented results of an experiment looking at the effect of manipulated vocal cues on implicit trustworthiness attributions. It was shown that high pitch and slow articulation rate increase participants' investments in the game, and that this effect is maintained over time. Also, participants invested more money with the female speaker than the male speaker. While the effect of high pitch was manifest from the first round of the game, the effect of rate and speaker gender became apparent after some time. Taken together with results of Experiment 2, these results suggest that people do pay implicit attention to vocal cues, and that these shape our perceptions of a speaker from the very first moment. These impressions are subsequently either reinforced or overturned in time, as we get to know the speaker through his/her actions. Also, some cues, such as articulation rate, start having an effect on implicit impressions only after some time. This again highlights the importance of using a methodology that allows to capture the process of formation and development of personality attributions. Finally, it seems that there is a particular sensitivity to vocal cues that are presumably extremely different from what one is used to, such as very slow speech rate. While a moderately slow rate increased perceptions of trustworthiness, this tendency stopped after an "uncanny" boundary was crossed (see the particularly slow speaker in Experiment 2), after which participants' discomfort was revealed through low investments (in Experiment 2) and low naturalness ratings (in the current experiment).

Together with Experiment 2, these results also suggest that pitch in particular plays a very important role in implicit trustworthiness perception. A higher pitch – relative to a speaker's average pitch – is also a cue to emotional expression, for example through smiling (Torre, 2014). Previous research has shown that smiling faces increase trustworthiness attributions, both explicitly and implicitly (Scharlemann et al., 2001). Since smiling can be heard from the speech signal alone (Tartter & Braun, 1994), it would be reasonable to assume that a "smiling voice" would increase trustworthiness judgments as well. However, no research so far has examined the impact of smiling while speaking on implicit trustworthiness attributions. Results of an experiment aimed at studying the effect of smiling voice through the investment game are presented in the next chapter.

7 Experiment 4 — Smiling voice

7.1 Introduction

The previous chapter dealt with how suprasegmental characteristics of the voice, in particular pitch and articulation rate, influence implicit trustworthiness judgments. Apart from providing the listener with information about the speaker's physical qualities (such as gender and age), though, these characteristics also contribute to conveying the emotional states that the speaker is in. Speaking very fast, for example, can be a sign of agitation or anxiety, while speaking slowly can inform the listener that the speaker is calm, or bored (Scherer, 1987). As will be discussed below, "halo effect" theories suggest that people displaying a positive emotion will be attributed other positive traits, too, for example trustworthiness. What is not clear is how a positive emotion expressed in the audio channel will mediate these halo effects. This chapter presents an account of an audible signal which is poignant with positive emotional expression – smiling – and shows results from an investment game where the smiling voice of a virtual agent is manipulated, in order to determine whether, and how, this expression influences implicit trust.

Emotional expressions are essential aspects of social interaction, so much that having difficulties in producing or perceiving them can be a sign of mental illness (Russell, Bachorowski & Fernández-Dols, 2003; C. G. Kohler, Turner, Gur & Gur, 2004). Even though emotions can be felt without being overtly expressed, and without being in an interactive context, it is believed that one of their original functions was essentially interpersonal and social, to alert others in a group about a change in the environment (Ekman, 1999). Thus, emotional displays help inform others about what actions to take (Van Kleef, De Dreu & Manstead, 2010; Klinnert, Campos, Sorce, Emde & Svejda, 1983). This hypothesis is supported by the fact that every human society has a concept of emotions and is able to recognize them (Ekman & Friesen, 1971; Ekman, 1992; Elfenbein & Ambady, 2002; Scherer & Wallbott, 1994; Bryant et al., 2016; Izard, 1969; Scherer, Banse & Wallbott, 2001). Nevertheless, there are differences in the way they are expressed (Brody, 2000; Matsumoto, 1993) and experienced (Rosaldo, 1984; Kitayama, Mesquita & Karasawa, 2006). Importantly, it has been argued that emotional expressivity – i.e. the capacity of displaying emotions at varying intensities – facilitates social communication, and is therefore a winning evolutionary strategy (Boone & Buck, 2003). Thus, individuals who are able to display and detect emotions accurately might be at an advantage in human societies.

Emotional expressivity is also linked to attractiveness and other positive traits (Sabatelli & Rubin, 1986; Reis et al., 1990), including cooperation and trustworthiness (Schug et al., 2010; Lount, 2010). Conversely, being able to detect trustworthy individuals might be a critical component in the evolution of cooperation (Russell et al., 2003; Boone & Buck, 2003). As a consequence, nonverbal expressions such as emotional displays might be “honest signals” of trustworthiness (Boone & Buck, 2003; Schug et al., 2010). However, while a significant amount of research has been conducted on trust as a rational behaviour, not much has been dedicated to study the influence of emotion on trust (Lount, 2010). Also, if displays of emotions play such a crucial role in the detection of trustworthiness, would emotions expressed in the auditory channel also influence trustworthiness perception?

One emotional expression that seems universal to humans and other animals is smiling (Mehu & Dunbar, 2008; van Hooff, 1972), which is produced by moving muscles in the mouth and eye regions (Eisenbarth & Alpers, 2011). As previously mentioned (see Section 5.4), a relatively small body size can be a sign of non-threat; for this reason, Ohala (1980) hypothesised that retracting the lips, which produces smiling in the lower part of the face, might have a become a sign of submission and non-threat because the corresponding sound made with a raised pitch might suggest a smaller body size. This facial expression is also similar to the "silent bared-teeth" display – which can still be observed in certain monkeys – which is a signal of submission (Preuschoft, 1992).

In human interaction, genuine or "Duchenne" smiles (Ekman, Davidson & Friesen, 1990) are mostly considered expressions of positive emotions such as happiness, joy, amusement or friendliness (K. J. Kohler, 2008; Ekman & Friesen, 1982; Thompson & Meltzer, 1964; Messinger, Fogel & Dickson, 2001). However, smiles can also be false, derisive or sad (Ekman & Friesen, 1982; Ambadar, Cohn & Reed, 2008).

In terms of person perception, displaying positive emotions is generally associated with positive trait attributions. For example, in a series of experiments with Chinese (Lau, 1982), Brazilian (Otta, Lira, Delevati, Cesar & Pires, 1994) and American participants (Reis et al., 1990; Mueser, Grau, Sussman & Rosen, 1984), it was found that pictures of smiling faces were rated higher in terms of attractiveness, friendliness and sincerity, among other traits, as compared to neutral faces. Looking specifically at trust, Elkins and Derrick (2013) found that participants rated a smiling Embodied Conversational Agent as more trustworthy than a non-smiling one. In an experiment using a trust game with three possible moves, Scharlemann et al. (2001) found that participants trusted counterparts who were represented as smiling photographs more than non-smiling ones. Also, participants in Krumhuber et al. (2007)'s study indicated that avatars displaying an authentic smile might be more cooperative in a trust game; subsequently, participants proceeded to invest more money with them in an actual one-shot game.

There are a few theories on the reason why displaying a positive emotion leads to the attribution of another positive trait, such as cooperativeness or trustworthiness. After findings that smiling faces elicited positive trait attributions, Lau (1982) argued that this could be due to a "halo effect": smiling is good, so people who smile must possess other good traits, too. Similarly, Penton-Voak, Pound, Little and Perrett (2006) concluded that people with facial features that elicit attributions of agreeableness may be treated as more trustworthy and may consequently develop more agreeable personality characteristics. This has been referred to as the "what is beautiful is good" stereotype (Dion et al., 1972; Reis et al., 1990). Additionally, since it has been pointed out that individuals who are particularly expressive might be less apt at disguising their emotions (DePaulo, Blank, Swaim & Hairfield, 1992), it has been suggested that emotional expressions might be a sign of commitment to cooperative behaviour (Schug et al., 2010; Frank, 1988). As a consequence, individuals who display emotions might be seen as less deceptive and more cooperative (Schug et al., 2010).

Another theory, the “Emotions as Social Information” (EASI) model (Van Kleef et al., 2010), mentions the importance of context. This model suggests that emotions are used to make sense of ambiguous situations, and that their effect depends on the situation in which the interaction takes place, specifically 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. On the contrary, displaying a negative emotion, such as anger, in a cooperative context will hinder future cooperative behaviours, and so on. Supporting this, Antos, De Melo, Gratch and Grosz (2011) found that, in a negotiation game, participants selected more often as partners computer agents that displayed emotions which were congruent with their actions. These agents were also perceived as more trustworthy than agents whose emotional expression and action strategy did not match, even though the strategy they had was the same. Thus, smiling might also have a differential effect based on the context: smiling in a cooperative situation might elicit further cooperation, while smiling in a competitive situation might elicit further non-cooperation.

When the visual channel is not present, people are able to detect emotions and emotional signals, such as smiling, from the voice alone. The occurrence of smiling while speaking has been given various names, for example "speech-smile" (K. J. Kohler, 2008; El Haddad, Cakmak, Moinet, Dupont & Dutoit, 2015), "smiled speech" (Émond & Laforest, 2013) or "smiling voice" (Pickering et al., 2009; Torre, 2014). As stated above, acoustically, smiling affects the vocal tract by shortening it (Shor, 1978), which contributes to raising the f0 and formant frequencies (K. J. Kohler, 2008; Tartter, 1980; Drahota, Costall & Reddy, 2008;

Fagel, 2009; Torre, 2014; Lasarcyk & Trouvain, 2008; Barthel & Quené, 2015). Several perceptual studies have demonstrated that the presence of smiling is audible in the speech signal alone (Aubergé & Cathiard, 2003; Torre, 2014; Drahota et al., 2008; Tartter & Braun, 1994), and is even detectable when the speech stimuli had been synthesised (Lasarcyk & Trouvain, 2008; El Haddad, Cakmak et al., 2015; El Haddad, Dupont, d'Alessandro & Dutoit, 2015).

Thus, given that smiling can be heard in the voice, and given that smiling might elicit positive feelings for the reasons discussed above, would a smiling voice elicit increased monetary investments in the investment game? And how would the smiling interact with a congruent-positive speaker's behaviour or with an incongruent-negative one? Van Kleef et al. (2010)'s EASI model would predict that smiling might increase trustworthiness in a congruent context (such as a *generous* behaviour) and decrease it in an incongruent context (such as a *mean* behaviour). Thus, a virtual player who displays a positive emotion through smiling, while returning more money to the participants than he/she invested might increase a feeling of mutual benefit and cooperation. On the contrary, a smiling virtual player who is consistently returning less money might increase the feeling that the virtual player is gaining at the participant's expense (Van Kleef et al., 2010; Lanzetta & Englis, 1989). Additionally, how would the presence or absence of smiling voice interact with other audible speaker characteristics, such as accent?

	Accent	**Smiling**	**Behaviour**
1	SSBE	Smiling	Generous
2	SSBE	Smiling	Mean
3	SSBE	Non-Smiling	Generous
4	SSBE	Non-Smiling	Mean
5	Birmingham	Smiling	Generous
6	Birmingham	Smiling	Mean
7	Birmingham	Non-Smiling	Generous
8	Birmingham	Non-Smiling	Mean

Table 6: Experimental conditions of the investment game. Accent and smiling were counterbalanced across participants so that each participant would play one game for each accent and smiling condition, while behaviour was counterbalanced so that all possible combinations (including playing e.g. two *mean* games) were allowed.

This chapter deals with the effect of vocal emotional cues on implicit judgments of trustworthiness, and with their interaction with other indexical and behavioural cues. It presents results of an investment game where participants played with virtual agents which were associated with different accents and smiling conditions. In particular, participants played with virtual agents which had either a Standard Southern British English or a Birmingham accent, which were either smiling or not, and which were behaving either *generously* or *meanly*. A summary of the experimental conditions can be found in Table 6.

7.2 Method

7.2.1 Participants

Participants were 110 native British English speakers (85 females, 24 males) aged 18-48 (median = 19, SD = 5.5). They were university students who received course credit for participation. Their geographic origins were reported as: southwest England (n = 61), southeast England (n = 27), Midlands (n = 9), Wales (n = 5), East Anglia (n = 3), northwest England (n = 2), northeast England (n = 1). One participant was excluded due to not being a native of the British Isles, and the data from another one were not recorded due to a technical error. Therefore, only the data of 108 participants were analysed.

7.2.2 Stimuli

Recordings were obtained from two female SSBE speakers and two Birmingham speakers. Each speaker was recorded in a sound-proof recording room reading two 20-sentence blocks each (see Appendix A. After they were recorded reading the sentences in their typical, "straight" voice, they were asked to re-read the sentences while sounding "amused". In order to elicit genuine amusement, the speakers were shown a selection of funny videos available on-line. These are common procedures used to elicit smiling voice (e.g. El Haddad, Cakmak et al., 2015). A researcher was in the same room of the speakers during the recordings, making sure that they were consistent while reading in the two conditions, and asking them to repeat the recordings in case they were not.

Apart from the presence of the researcher, the fact that the smiling recordings sounded more "amused" than the typical ones was validated by ratings in the post-game questionnaire that participants completed (see section 7.2.3). Furthermore, f0 measurements of the two sets of recordings, obtained with Praat (Boersma & Weenink, 2017), show that the smiling recordings had a statistically higher f0 than the non-smiling ones (mean smiling f0 = 206 Hz, mean non-smiling f0 = 191 Hz; $t(317.23) = 10.38$, $p < .001$), as shown in Figure 19. As mentioned before, raised f0 is a typical characteristic of smiling voice (K. J. Kohler, 2008; Tartter, 1980; Drahota et al., 2008; Fagel, 2009; Torre, 2014; Lasarcyk & Trouvain, 2008; Barthel & Quené, 2015).

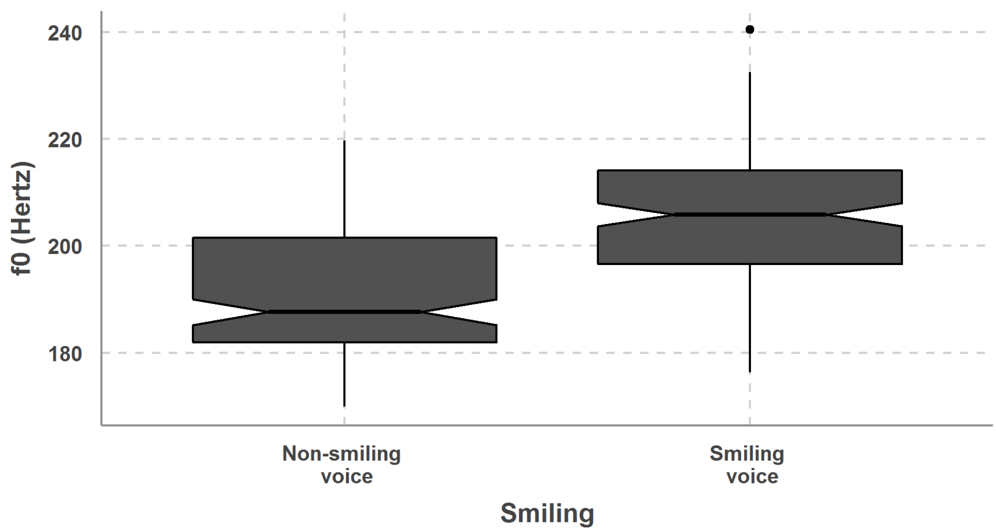


Figure 19: Average f0 in the smiling and typical recordings.

The articulation rate was also significantly faster in the smiling condition (mean smiling rate = 4.36 syllables/second, mean non-smiling rate = 4.21 syllables/second; $t(316.92) = 2.37$, $p = .018$), although this feature has rarely been investigated in studies on the acoustics of smiling voice.

7.2.3 Procedure

The experiment followed the same procedure as the previous three experiments (see chapters 4, 5, and 6). Each participant engaged in two games, one for each accent, smiling and behaviour condition (*generous*/*mean*), with a different block of 20 utterances heard for each version of the game. The experiment was counterbalanced in a 2 (accent: SSBE or Birmingham) by 2 (voice: smiling or non-smiling) within subject design. Furthermore, the behaviour (*generous* or *mean*) condition was counterbalanced across participants in all possible combinations, so that some participants played two consecutive games with the same behaviour condition. The sentence block and speaker number were randomized across participants. The game rounds proceeded as in the previous experiments.

After completing the first game, participants answered a few questions about the virtual player they had just played with, on a 7-item Likert scale (1 = strongly disagree, 7 = strongly agree). The questions were taken from Rau et al. (2009): "This person was sincere"; "This person was interested in talking with me"; "This person wanted me to trust her"; "This person was honest in communicating with me". Again, these questions were intended to act as a safety check that participants had explicitly understood the behaviour of the virtual player and, consequently, how the game worked. The same questions were asked again at the end of the second game. After completing the second game and after answering these questions related to the second speaker they heard, participants were played 10 utterances from one of the speakers they had just heard, chosen at random, and asked to rate them in terms of how happy they sounded, on a 7-item Likert scale (1 = very unhappy, 7 = very happy). Then, participants were asked to rate how much they liked the voice they had just heard, on a 7-item Likert scale (1 = strongly dislike, 7 = strongly like). The same questions (happy-sounding and voice-liking) were then asked about the other speaker they heard in the game. This was done to confirm that the smiling and amusement could indeed be heard in the recorded voices. They were then asked to state the reason behind their answer. Then they were asked to identify the accent of the speakers they heard. Like in Experiment 2, they were given a list of possible accents to choose from, including the two correct answers, 5 incorrect answers, and "Other" and "Don't know". The option for the SSBE accent was described as "Standard English". Finally, participants completed a short background questionnaire, where they were asked their age, gender, city of origin, and what accent they spoke. As a final question, participants were also asked what accent they would like a robot to have; this is discussed in full later, in Section 8.5. The total duration of the experiment was approximately 20 minutes.

7.3 Results

7.3.1 Investment game

Unfortunately, due to a technical error which was only discovered at the data analysis stage, participants only heard one Birmingham speaker, "Birmingham 2" in the *generous* condition, and the other, "Birmingham 1" in the *mean* condition. The SSBE speakers were correctly played in both conditions. This leads to a confound between the Birmingham speakers and the behaviour condition. In order to deal with this, the full dataset was analysed as if there was only one speaker for each accent, as was the case in Experiment 1 – thus, the speaker variable was not added to the main model. As a secondary analysis, a model was also fitted to a subset of the original data without the Birmingham condition, in order to, at least partially, examine the SSBE speakers' idiosyncrasies in relation to the smiling conditions.

To determine the effects of game behaviour and voice on investments, a mixed-effects linear model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC (Aho et al., 2014). Investment was the dependent variable; behaviour, game turn, accent and smiling were predictors; subject and sentence block – i.e. the two different sets of 20 recorded sentences each – were random factors.

There was a main effect of behaviour, with an average investment of £5.63 to the *generous* virtual player and of £3.31 to the *mean* virtual player ($\chi^2(1) = 1135$, $p < .001$). There was also a main effect of game turn ($\chi^2(1) = 276.37$, $p < .001$), with higher overall investments in the second half of the game, and a significant interaction between behaviour and game turn ($\chi^2(1) = 27.32, p < .001$): as in previous experiments, investments increase in the *generous* condition and decrease initially, and then increase, in the *mean* condition as the game progresses (Figure 20).

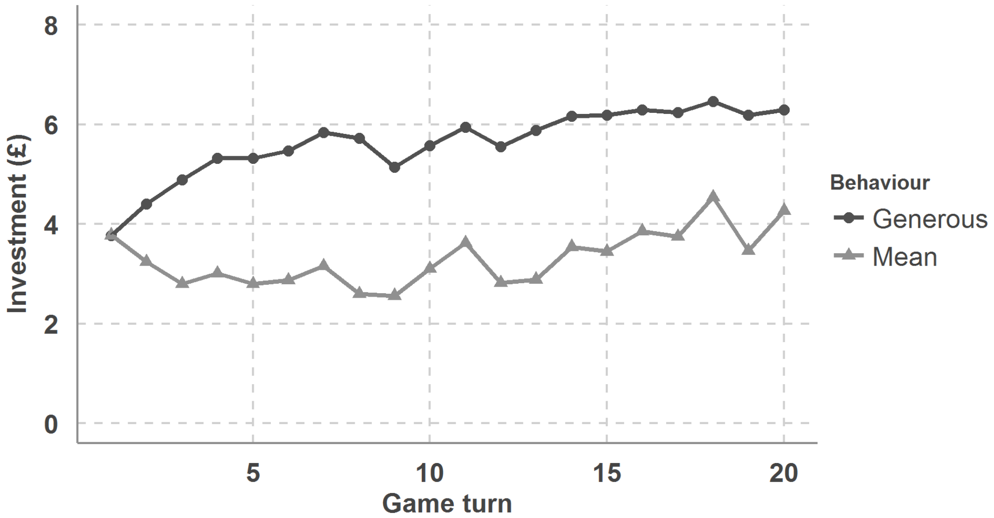


Figure 20: Average investments in the *generous* and *mean* conditions.

There was also a main effect of smiling, with participants investing more in the smiling voice than the non-smiling voice (mean overall investments to smiling voice = £4.58, to non-smiling voice = £4.38; $\chi^2(1) = 13.65, p < .001$). There was no main effect of accent, however.

There was a significant three-way interaction between behaviour, accent and smiling ($\chi^2(5) = 34.91, p < .001$, Figure 21). An ANOVA was used to analyse this interaction, splitting the data in the two behaviour conditions. In the *generous* condition, there was no effect of smiling ($F(1, 2176) = 2.78, p = .096$), but there was an effect of accent ($F(1, 2176) = 13.75, p < .001$), with higher investments to the SSBE virtual player (mean = £5.81) than to the Birmingham virtual player (mean = £5.46). In the *mean* condition, instead, there was a main effect of smiling ($F(1, 2136) = 4.37, p = .036$), with higher investments in the smiling (mean = £3.41) than the non-smiling condition (mean = £3.20). There was also a main effect of accent ($F(1, 2136) = 25.92, p < .001$), with higher investments in the Birmingham virtual player (mean = £3.56) than the SSBE-accented one (mean = £3.05). There was also a significant interaction between smiling and accent ($F(1, 2136) = 4.79, p = .029$): a post-hoc comparison using the Tukey HSD test showed that investments to the non-smiling Birmingham speaker were significantly higher than to the non-smiling SSBE speaker (mean = £3.56 and £2.84 respectively, p < .001), that investments to the smiling Birmingham speaker were significantly higher than to the non-smiling SSBE speaker (mean = £3.56 and £2.84 respectively, $p < .001$), and that investments to the smiling SSBE speaker were higher than to the non-smiling SSBE speaker (mean = £3.27 and £2.84 respectively, $p = .012$).

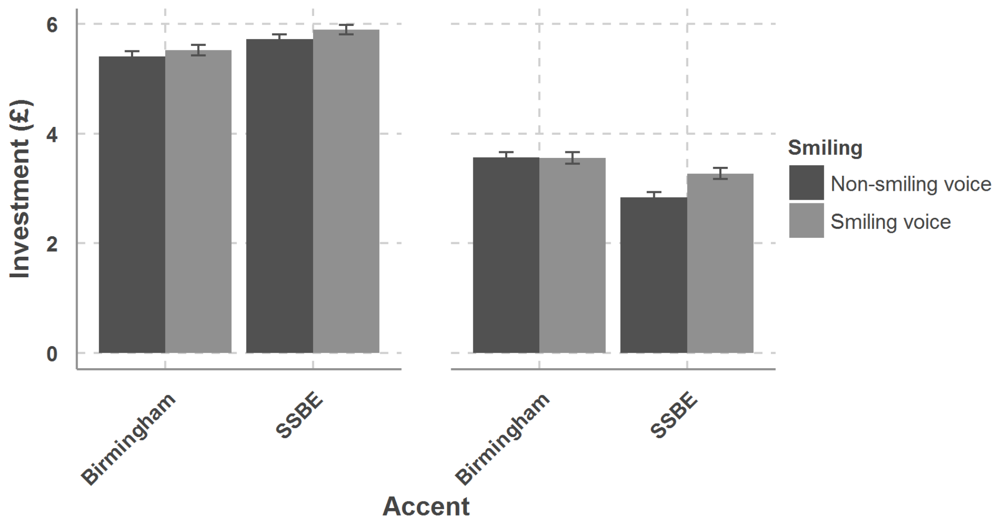


Figure 21: Average investments in the two accent and smiling conditions, in the *generous* (left) and *mean* (right) behaviour conditions.

There was also a significant three-way interaction between behaviour, accent and game turn ($\chi^2(2) = 17.53, p < .001$, Figure 22). A post-hoc ANOVA showed that, in the *generous* condition, there was a main effect of game turn ($F(1, 2176) = 168.87, p < .001$), with investments increasing over time. There was also a main effect of accent ($F(1, 2176) = 15.16, p < .001$), with higher investments to the SSBE speaker, as mentioned before. In the *mean* condition, there was an effect of game turn, with investments decreasing in the first half of the game and then increasing in the second half of the game ($F(1, 2136) =$

39.09, $p < .001$). There was also a main effect of accent ($F(1, 2136) = 26.07$, $p < .001$), with higher investments to the Birmingham speaker, as mentioned before.

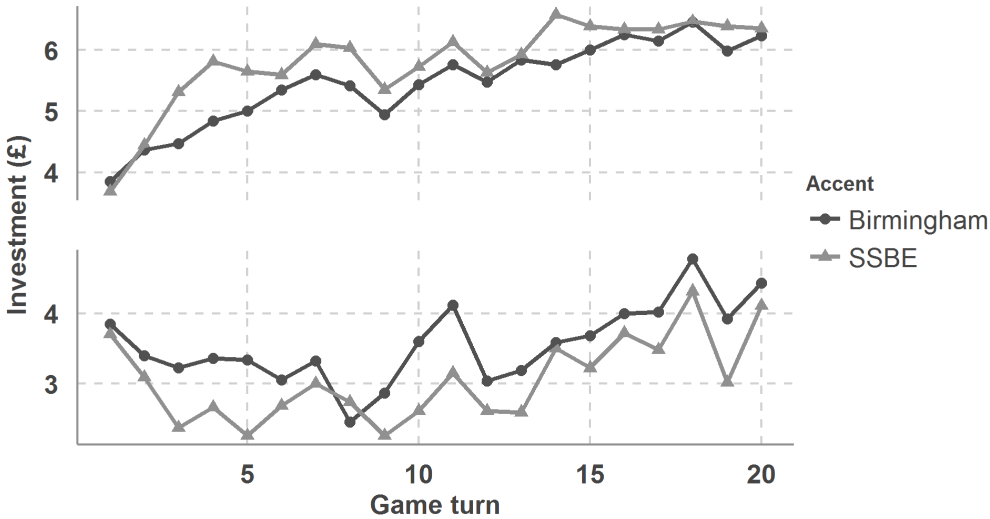


Figure 22: Average investments in the two accents in the *generous* (top) and *mean* (bottom) conditions over time.

As a secondary analysis, a mixed-effects linear model was fitted to a subset of the data not containing the Birmingham condition. Investment was the dependent variable, behaviour, game turn, smiling and (SSBE) speaker identity were predictors, and participants and sentence blocks were random factors. Due to the design of the experiment, this subset of the original data was completely between-subjects, since participants heard a different accent in each of the two games they played.

There was a main effect of game turn ($\chi^2(1) = 149.36$, $p < .001$), with investments increasing over time. There was also a main effect of behaviour ($\chi^2(1) = 81.89$, $p < .001$), with higher investments in the *generous* condition (mean = £5.81) than the *mean* condition (mean = £3.06). There was also an interaction between behaviour and game turn ($\chi^2(1) = 11.96$, $p < .001$): as shown in Figure 23, investments in the *generous* condition increase more steadily at first, and then reach a sort of plateau. In the *mean* condition instead, investments decrease initially and then increase in the second half of the game.

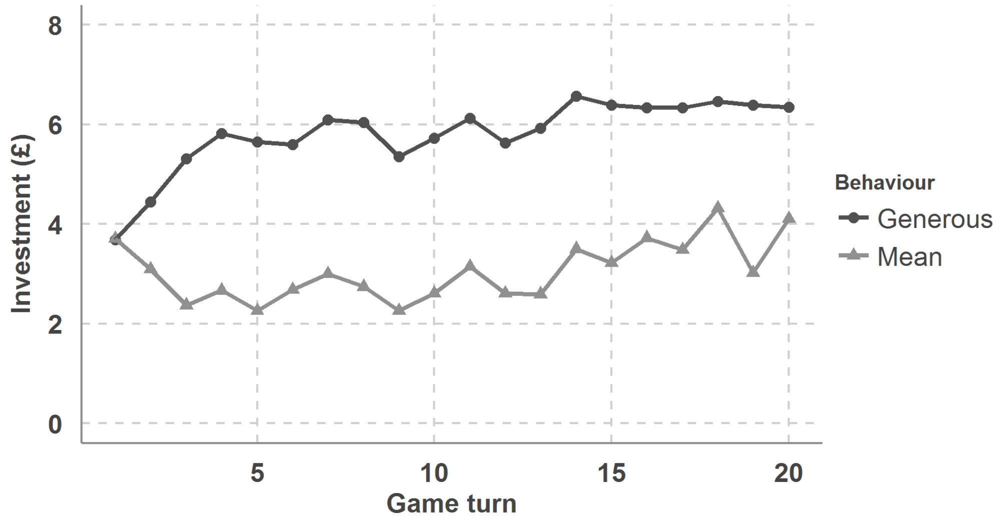


Figure 23: Average investments in the *generous* and *mean* conditions in the subsetted data.

There was also a main effect of speaker identity ($\chi^2(1) = 4.74$, $p = .029$), with higher overall investments to speaker "SSBE 1" (mean = £4.76) than to speaker "SSBE 2" (mean = £4.11). There was also an interaction between speaker and game turn ($\chi^2(1) = 29.76$, $p < .001$): as shown in Figure 24, initial higher investments to speaker "SSBE 2" are overturned after the first 3 rounds of the game, and higher investments to speaker "SSBE 1" are maintained throughout the rest of the game.

There was no effect of smiling, and no interactions.

7.3.2 Questionnaires

Explicit honesty and sincerity evaluations

The numerical answers to the 4 questions about the perceived trustworthiness of the speaker were the dependent measure in a mixed-effects model with smiling and accent as predictors and participant as random factor. There was a main effect of behaviour ($\chi^2(1) = 134.58$, $p < .001$): participants overall gave higher ratings to the speakers in the

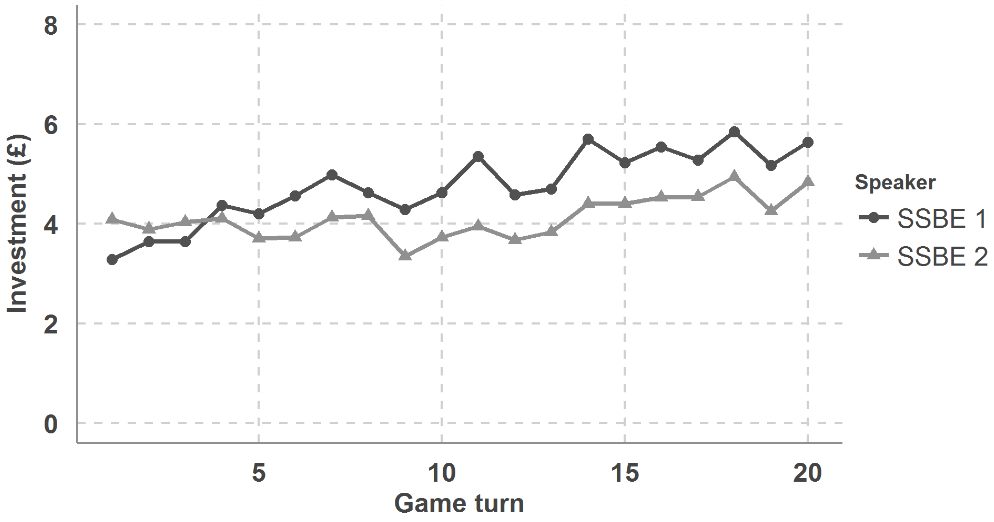


Figure 24: Average investments in the two SSBE speakers.

generous condition than in the *mean* condition (mean ratings = 5.13 and 3.55, respectively). This, together with the result that participants invested more money with the *generous* virtual players than the *mean* players, again confirms that they understood how the virtual player was behaving and reacted accordingly. There was also a main effect of smiling ($\chi^2(1) = 16.69$, $p < .001$): participants gave higher ratings to the smiling virtual player than to the non-smiling virtual player (mean ratings = 4.61 and 4.09, respectively). There was no interaction between smiling and behaviour, so this effect of smiling seems to be constant across behaviour condition. That is to say, no matter how a smiling speaker behaved in the game, it was still perceived as more honest and sincere than the non-smiling speaker. Finally, there was a marginally significant effect of accent ($\chi^2(1) = 3.80$, $p = .051$), suggesting that participants tended to rate the Birmingham speaker as more honest and sincere than the SSBE one (mean ratings = 4.47 and 4.22, respectively).

Happiness rating

A mixed-effects model with happiness rating as dependent variable, smiling and accent as predictors, and participant and sentence block as random factors was fitted to the happiness questionnaire data. There was a main effect of the smiling condition ($\chi^2(1) = 1360.8$, $p < .001$), with participants rating the smiling voice samples as sounding happier than the non-smiling voice samples (mean rating for smiling voice = 4.9, mean rating for non-smiling voice = 3.44). There was also a main effect of accent ($\chi^2(1) = 10.21$, $p = .0014$), with participants rating the Birmingham speakers as happier than the SSBE speakers (mean = 4.2 and 4.12, respectively). There was also an interaction between smiling and accent ($\chi^2(1) = 5.78$, $p = .016$): as shown in Figure 25, the Birmingham speakers sounds happier than the SSBE speakers in the non-smiling condition. This might be due to inherent speaker characteristics such as pitch or pitch range.

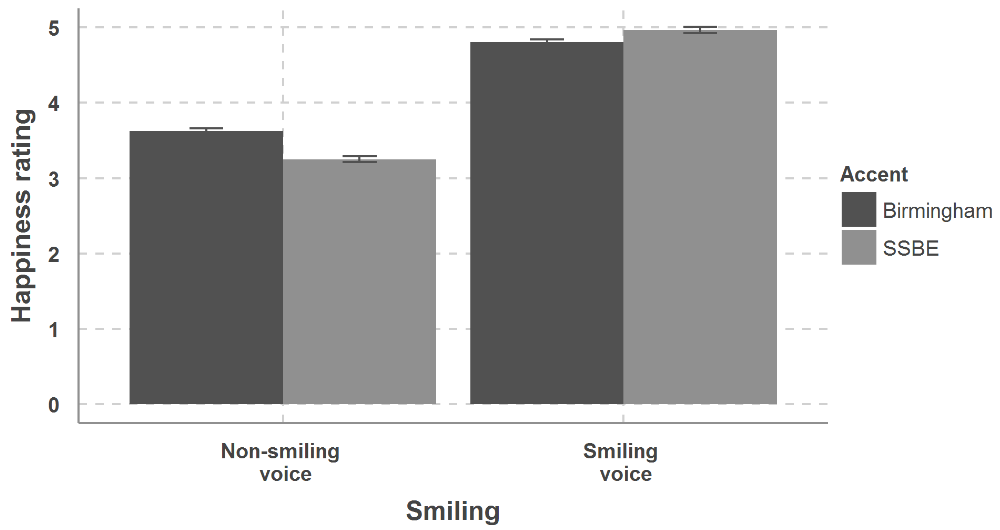


Figure 25: Happiness ratings in the two accents and smiling conditions.

This hypothesis is validated by the fact that a second model, built with smiling and speaker identity as predictors, resulted in main effects of smiling ($\chi^2(1) = 1360.8$, $p < .001$) and of speaker identity ($\chi^2(3) = 117.59$, $p < .001$), but in no interaction between speaker and smiling. A comparison between the model with accent and the model with speaker shows that the latter fits the data better ($\chi^2(1) = 101.6$, $p < .001$).

A post-hoc Tukey HSD test revealed that speaker "Birmingham 1" sounded significantly happier than speakers "Birmingham 2", "SSBE 1" and "SSBE 2" ($p < .001$, $p < .001$ and $p = .003$, respectively), and that speaker "SSBE 2" sounded significantly happier than speakers "Birmingham 2" and "SSBE 1" (all $p < .001$), as shown in Figure 26.

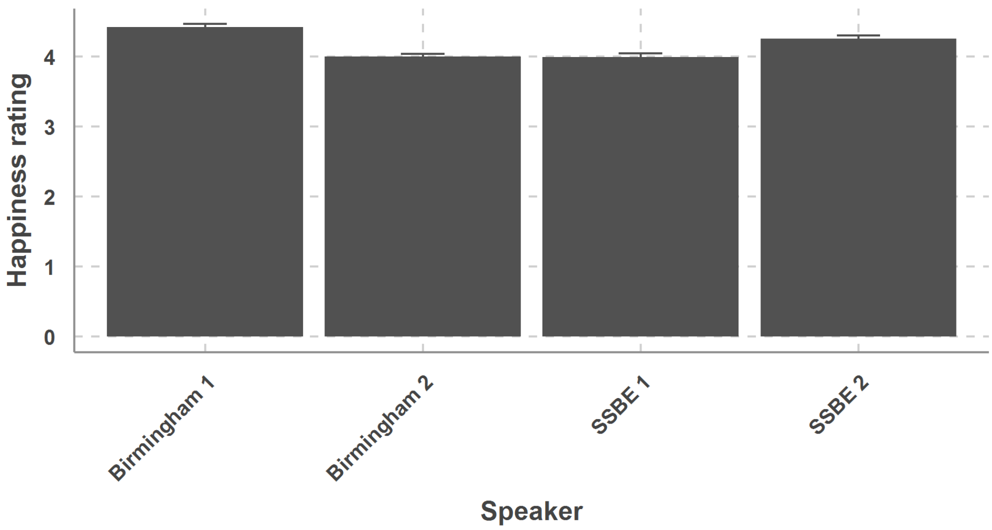


Figure 26: Happiness ratings of the 4 speakers, collapsed across smiling conditions.

Although individual speaker characteristics were not controlled for in this experiment, looking at them post-hoc might help elucidate on these different speaker ratings. Of the 4 speakers recorded, "Birmingham 1" had the highest f0 in both smiling and non-smiling recordings, as shown in Figure 27.

A one-way ANOVA showed that the f0 differed significantly among the 4 speakers ($F(3, 316) = 82.9$, $p < .001$). A post-hoc Tukey HSD test revealed that "Birmingham 1"'s f0 was significantly higher than "SSBE 1"'s, "SSBE 2"'s and "Birmingham 2"'s (all $p < .001$); also, both "SSBE 1" and "SSBE 2" had a significantly higher f0 than "Birmingham 2" (all $p < .001$).

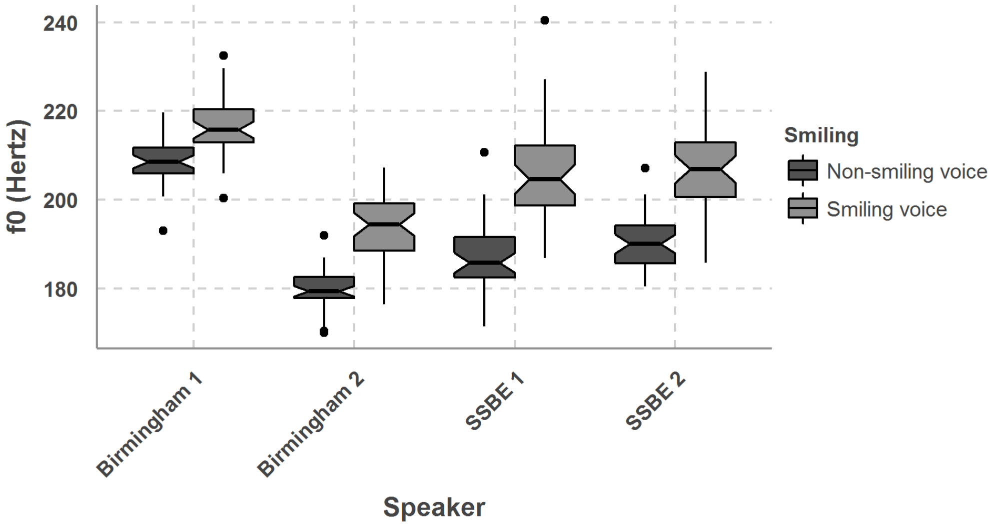


Figure 27: Average f0 of the 4 speakers.

A one-way ANOVA also showed that articulation rate differed significantly among the speakers ($F(3, 316) = 2.64$, $p = .049$). A post-hoc Tukey HSD test revealed that the articulation rate of speaker "Birmingham 2" was faster than the one of speaker "SSBE 1" (mean = 4.37 and 4.14 syllables/second, respectively; $p = .49$). The articulation rate of all the speakers is shown in Figure 28.

Thus, speaker "Birmingham 1" had the highest f0, so it is possible that this particular feature influenced ratings of happiness.

Accent identification

The SSBE accent was correctly identified by 66 out of 108 participants (61%), and 45 correctly identified the Birmingham accent (42%). Given that participants could choose out of 9 options, in both cases performance was above chance level (cut-off = 11.2%).

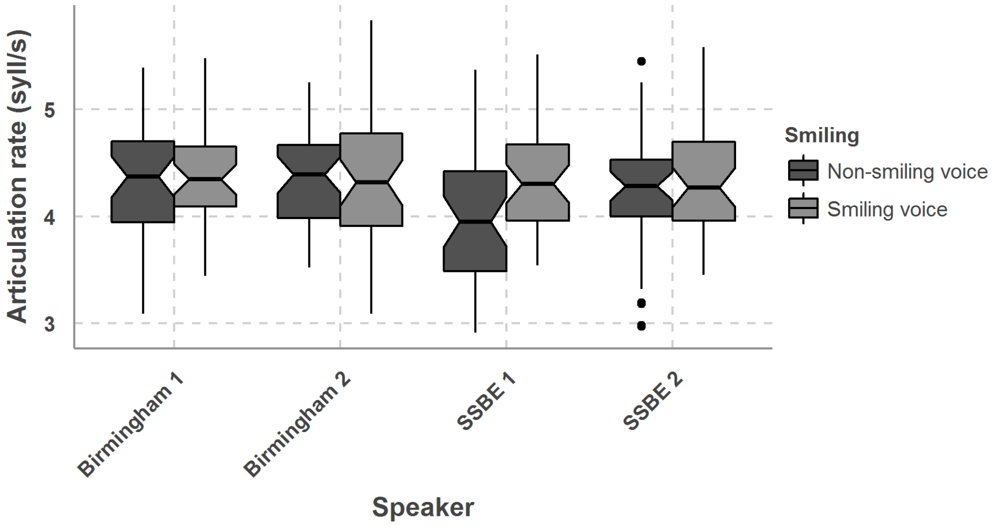


Figure 28: Average rate of the 4 speakers.

Voice liking

A mixed-effects linear model was fitted to the voice liking questionnaire. The rating was the dependent variable; smiling, accent and speaker were the predictors; participant and sentence block were random factors. There was a main effect of smiling ($\chi^2(1) = 63.61$, $p < .001$), and no effect of accent or speaker: participants liked the smiling voices significantly more than the non-smiling ones, independent of who was speaking.

7.4 Discussion

This experiment investigated whether positive vocal emotional expression, in the form of smiling voice, influenced implicit trust attributions, and whether it interacted with other vocal characteristics (such as accent) and with speaker's behaviour. Behaviour, game turn and smiling all influenced monetary investments in the game, while accent did not. In particular, participants invested more money in the *generous* condition, investments increased over time, and there were higher investments to the smiling virtual players.

The effect of smiling in the game and in the questionnaire adds to previous studies on emotional expression, showing that the display of a positive emotion increases trust and likeability also in the verbal channel (Scharlemann et al., 2001; Krumhuber et al., 2007; Penton-Voak et al., 2006; Lau, 1982). Furthermore, this result extends previous findings to implicit emotion processing, since the majority of studies on the association between positive emotion expression and positive traits used only explicit methods such as questionnaires (Otta et al., 1994; Reis et al., 1990).

Smiling was a main predictor of investments overall. That is to say, while participants still invested more or less money depending on the other player's trustworthiness (or lack thereof), they also overall invested more money in the smiling players. This contrasts with Van Kleef et al. (2010)'s EASI model, according to which the display of a positive emotion in an incongruent context (such as the *mean* behaviour condition) should elicit uncooperative behaviours. Quite on the contrary, in the current experiment smiling had a somewhat reduced effect in the congruent, *generous* condition – as shown by the post-hoc interaction analysis – and the strongest effect in the *mean* condition. While Van Kleef et al. (2010) listed social dilemma tasks based on Prisoner's Dilemma among possible competitive situations, it is possible that actual participants in an iterated game might view it as an essentially cooperative task. Game theoretical implications and speculations about this, and similar results from the other experiments, are discussed in Section 9.5.

If participants are reluctant to give up on cooperation — as shown by the fact that investments increase in the second half of the game in the *mean* condition — they might be even more reluctant to give up on partners who seem to encourage them to cooperate, with their positive emotional expression. Krumhuber et al. (2007) found that people explicitly and implicitly trusted smiling faces more than neutral faces, regardless of the sincerity of their smile, and that genuine smiles were trusted more than fake smiles. Similarly, Reed, Zeglen and Schmidt (2012) found that people displaying either Duchenne or non-Duchenne smiles were more likely to cooperate in a one-shot investment game. Thus, displaying an emotion, even a feigned one, might be preferred to not displaying any emotion at all, hence the increased investments to the *mean* smiling players.

Another possibility is that participants might have been "infected" by the positive emotion displayed by the smiling virtual players. In fact, emotional expressions can evoke affective reactions in observers, which may subsequently influence their behaviour (Hatfield, Cacioppo & Rapson, 1994), and this "emotional contagion" might well be transmitted through the auditory channel as well (see Magnée, Stekelenburg, Kemner & de Gelder, 2007). If this is the case, participants might have trusted smiling players more because, feeling a positive emotion themselves, this might have prompted them to behave in a cooperative manner (Schug et al., 2010).

The effect that was found for smiling also suggests that, in the absence of visual information, the audio equivalent of a Duchenne smile might act as a "honest signal" of cooperation. As mentioned before, Duchenne smiles are smiles describing genuine joy or amusement (K. J. Kohler, 2008; Ekman & Friesen, 1982; Thompson & Meltzer, 1964). Traditionally, in the visual domain they can be distinguished from other types of smiles because they involve the contraction of the "orbicularis oculi" muscle, which is a movement that is notoriously difficult to fake (Ekman & Friesen, 1982; Schug et al., 2010). Obviously, in the auditory channel it is not possible to detect a genuine smiling voice from this muscular movement. However, it is possible that a smiling voice which sounds amused might well be the auditory equivalent of a Duchenne smile. As participants indicated that the smiling voices used in this study did sound amused, there is reason to believe that the expression of amusement in the speech signal might lead listeners to believe that the speaker can be trusted. Unfortunately, no video recordings were taken during the audio recording of the speakers used in this experiment. If a similar study were to be replicated, the actual facial expression of the speakers should be manipulated in order to determine whether the amused expression in the voice corresponded to an actual Duchenne smile in the face of the speakers.

The two accents used in the current experiment, SSBE and Birmingham, were also used in Experiment 2. In Experiment 2, these two particular accents had the strongest effect on investments, with participants investing the highest amounts of money in the SSBE- and Birmingham-accented virtual player in the *generous* condition, and the lowest amounts to the Birmingham-accented player in the *mean* condition (see section 5.3.1). In the current experiment, there was no main effect of accent, but there was an interaction between accent, behaviour and smiling (Figure 21): in the *generous* condition, participants invested more money in the SSBE-accented players, independent of the smiling condition, while in the *mean* condition participants invested more money with the smiling virtual players and with the Birmingham-accented players. Also, in the *mean* condition participants trusted a smiling SSBE speaker more than a non-smiling one. Consistent with results shown in previous chapters, when the situation is beneficial for the participant (*generous* condition), stereotypical information that participants might obtain on the basis of voice influences investment decisions. The standard-accented virtual player is behaving as expected, and is therefore trusted more; the regional-accented virtual player is not behaving as expected, and is therefore trusted less. Within these accents, the presence of a display of a positive emotion increases the feeling of mutual benefit, and leads to a slightly higher trusting behaviour. On the other hand, when the situation is not working out well (*mean* condition), positive emotional cues seem to somehow mitigate the fact that the standard-accented virtual player is not behaving as expected. This supports the results of Experiment 1, where participants invested more money in the SSBE-accented player in the *generous* condition, and more in the Liverpool-accented player in the *mean* condition. Again, people might be more trusting towards an accented-speaker who behaves in an expected way, which is congruent with a listener's preconception; on the contrary, accented players who behave incongruously from what participants might expect of them, given their preconceptions, might be penalized by investing less money in them.

Still, the overall main effect of smiling suggests that vocal emotional expressions play a bigger role than speaker's accent in eliciting perceptions of trustworthiness. This can be interpreted as evidence of the fact that emotional expressions might be a "honest signal" of cooperativeness (Boone & Buck, 2003), while indexical information such as a speaker's accent might be integrated in the process of impression formation after the emotional expressivity has already influenced it. Thus, positive emotional display might override indexical information when it comes to implicit impression formation.

This experiment presents only incomplete evidence regarding individual speaker's cues. From the subset of game data, it was possible to evince that speaker identity played a role in implicit trust attributions, since participants invested more money in one of the two SSBE speakers in particular. Also, participants rated one of the two Birmingham speakers as particularly happy-sounding. This speaker happened to have the highest f0 in both her smiling and neutral speech. A high f0 positively influenced trust in Experiments 2 and 3, so it is possible that, if the game data had been complete, participants would have invested more money in this particular speaker as well. In the absence of such data, it is only possible to reiterate that impression formation goes beyond indexical information from the voice, and that other information which is transmitted in an individual's voice, such as their emotional state, plays a role in it as well.

7.5 Conclusion

This chapter presented results of an experiment looking at the effect of emotional cues and speaker's accent on implicit trustworthiness attributions. A voice showing signs of a positive emotion – in the form of smiling voice – consistently increased participants' investments in the game. This effect was constant over time and across the two behaviour conditions. That is to say, contrary to predictions stated above, people implicitly attribute trustworthiness to speakers showing a positive emotion in their voice, even when their behaviour contradicts this impression. On the other hand, while speaker's accent was not a main predictor of investments, its interaction with speaker's behaviour and game turn suggests that speakers judged initially more reliable may be more severely discredited and penalized if their behaviour is at odds with first impressions. Similarly to Experiment 1, speaker's accent affected how participants reacted to negative behaviour. The SSBE speaker initially attracted higher investment, but showed a greater drop in investment than the regional speaker (Liverpool in Experiment 1, Birmingham here) once the negative pattern of returns became evident. Within the *generous* condition, participants apparently trusted the standard-accented player more. Instead, in the *mean* condition, participants invested less in the SSBE-accented player.

The experiments presented so far have shown that different vocal cues influence implicit trustworthiness attributions which are only partially mediated by the simulated speaker's behaviour. In fact, if participants only took into account the virtual player's behaviour when deciding how much money to invest, we would see no difference in the relative investments to the different voices within the two simulated behaviours. However, within this thesis it has been shown that this is not the case.

This raises important issues in the context of voice-mediated Human-Machine Interaction. As such interactions become more and more common in our everyday life – vocal personal assistants like Siri, or robotic elderly care systems, to name a few – it becomes essential to build voices for these machines that elicit feelings of safety and trust. Furthermore, in order to make these interactions as smooth and natural as possible, it is necessary to design such systems' voices based on the context where the systems will be used. For example, vocal assistants are essentially just that: a voice. This voice might be studied in isolation, since it is unlikely that the form of the device that "speaks" it interacts or interferes with it. On the other hand, machines such as robots have a body that can generally move as well, and this is likely to interact with the type of voice that users might feel more comfortable with. For example, both the physical appearance and the voice of elderly care robots might be designed to elicit feelings of security and knowledge. On the other hand, robots intended to aid children's learning might be designed to resemble children's peers or pets, and so on. However, while a lot of attention has been dedicated to the physical appearance of robots, not a lot of research has been conducted on robotic voices, or on how they interact with the body they are attached to. The next chapter presents results of an experiment in Human-Robot Interaction aimed at untangling this potential interaction between a voice and its body.

8 Experiment 5 — Human-Robot Interaction

8.1 Introduction[1]

Anthropomorphism — the attribution of human-like physical features or mental states to other agents (Epley, Waytz & Cacioppo, 2007; Złotowski, Strasser & Bartneck, 2014; Lemaignan, Fink & Dillenbourg, 2014) — is an important topic in Human-Robot Interaction (HRI) research. Anthropomorphism derives specifically from the human need to control and predict, and unpredictability increases the tendency to anthropomorphise nonhuman agents (Waytz et al., 2010). In fact, anthropomorphism increases for agents whose morphology resembles that of humans, and conversely people tend to project their beliefs and desires on human-like stimuli (Waytz et al., 2010), and to expect more human-like machines to behave similarly to humans (B. R. Duffy & Joue, 2004). Anthropomorphism also might change with experience, as previous research suggests (Fussell, Kiesler, Setlock & Yew, 2008; Lemaignan et al., 2014). For example, in Fussell et al. (2008), participants rated a robot as possessing traits, moods and feelings more after they interacted with a robot, than after they simply imagined one. Thus, interacting with a robot increases its anthropomorphism. In outlining a formal model of anthropomorphism, Lemaignan et al. (2014) add that, apart from an interaction, the interaction context also influences the dynamics of anthropomorphism.

Stemming from this, since humans are highly skilled at interacting with other humans, it can be argued that they will find it easier to interact with human-like machines (Krach et al., 2008). As previously mentioned, a changing society will need humans to interact successfully with

[1] This experiment was run in collaboration with Ms. Debora Zanatto of the School of Computing, Electronics and Mathematics; her contribution is the robot's joint attention and the post-game questionnaires, while I worked on the voice and the investment game. For completeness, this chapter reports the full analyses and procedure of the experiment.

machines (Section 2.2.4), and machine designers will need to ensure that users do not feel threatened, or unsafe, about interacting with them. Thus, it becomes critically important, when it comes to designing robots, to understand how anthropomorphic their design should be in its various components. Too much human-likeness, in fact, can result in the negative Uncanny Valley effect (Mori, 1970), evoking feelings of disgust and eeriness. For example, androids have striking resemblance with human appearance (Minato, Shimada, Ishiguro & Itakura, 2004), but, as soon as they start to move, or speak, their machine qualities are immediately evident (Becker-Asano, Ogawa & Nishio, 2010). Such a mismatch between human appearance and machine movement affects people at the neural level, with fMRI data showing that androids with machine-like movements evoked larger prediction errors in the Action Perception System (Saygin, Chaminade, Ishiguro, Driver & Frith, 2012).

As previously discussed (Section 1.1), human-human interactions heavily rely on trust, and working robot-human relationship need to be based on trust as well (Hancock et al., 2011; Groom & Nass, 2007). On the one hand, humans need to trust a robot to not hurt them, to carry out its task, to function properly; on the other hand, the robots need to show that they are trustworthy. Since voice can act as a "honest signal" of trustworthiness (see Section 2.1.3), robot voices should be designed in order to elicit trustworthiness as well.

Previous research showed that being able to speak (either with a synthetic or a natural voice) was enough for a robot to be treated as a competent agent (Sims et al., 2009). Also, Nass et al. (1994) demonstrated that people apply to machine voices the same conversational and interactional rules that they would use with a human, and that the sound of the voice itself is enough to make personality judgments, independent of the speech content (Nass & Lee, 2001). Nass and Brave (2005) also found that male and female interactants reacted in the same way to natural and synthetic voices in a choice-making task, provided the voice corresponded to their gender. Arguably, different-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 and c) elicit personality attributions congruent with their task. For example, a big factory robot would appear rather strange if it had a very high-pitched voice. Or, children robot tutors meant to act as children's peers should have child-like voices.

Regarding personality traits attributions, previous literature is contradictory. For example, Shneiderman and Plaisant (2004) claimed that machines exhibiting human characteristics such as emotions or personality traits could confuse or mislead human interactants. Contrasting this, from an educational psychology point of view, it has also been found that virtual agents with a pre-recorded human voice led to participants scoring higher in a learning test than participants who heard a synthetic voice, while there were no differences in the perceived difficulty of the learning task (Atkinson, Mayer & Merrill, 2005). Finally, Mitchell et al. (2011) argued that incongruence in the human-likeness of a character's face and voice can elicit feelings of eeriness. The majority of social robots in use nowadays are not perfectly human-like, but rather their morphology resembles that of a human in some way or another. For example, the Nao robot is a biped, has two arms, and a head with a resemblance of a mouth and eyes (Figure 29). When it comes to designing such a robot's voice, the question of whether to use a natural or synthetic voice also arises. Would people prefer a "congruent" robot, which approaches human-ness in all its characteristics, but does not quite reach it? Or would people prefer an "incongruent" robot, with a pre-recorded human voice, for the sake of clarity and familiarity (Tamagawa, Watson, Kuo, MacDonald & Broadbent, 2011)? Some of the differences in results from previous studies are likely due to using different tasks and application domains. In general, however, we can argue that, given that people tend to attribute traits to computers and robots as if they were human agents (Nass & Lee, 2001; Nass et al., 1995; Walters, Syrdal, Dautenhahn et al., 2008), that people respond to robots as if they had a personality (Lee et al., 2006), and that people attribute traits to speakers based on very fine speech characteristics (see results from Experiments 2–3), there is reason to assume that they will attribute traits — e.g. of trustworthiness — to robots speaking in different voices as well. In the experiment presented in this Chapter, the voice and attention behaviour of a Nao robot were manipulated in order to be more or less anthropomorphic, and implicit trust attributions were measured using an investment game.

Figure 29: Softbank Robotics's Nao robot.

Another feature that has often been shown to influence anthropomorphism is joint attention. Interlocutors monitor the gaze of their conversation partners to establish joint attention (Moore & Dunham, 2014; Mason, Tatkow & Macrae, 2005), which also influences participants' decisions in HRI tasks (Mutlu, Shiwa, Kanda, Ishiguro & Hagita, 2009; Staudte & Crocker, 2011). By briefly moving their eyes, robots can affect the decision making process, even when participants do not report seeing those cues (Admoni, Bank, Tan, Toneva & Scassellati, 2011). Gaze and joint attention also influence person perception, including in terms of trust (Bayliss & Tipper, 2006; Staudte & Crocker, 2011; Mason et al., 2005). Previous studies on the topic come to diverging conclusions, however. For example, Bruce, Nourbakhsh and Simmons (2002) found that attention did not increase the participants' interest in interacting with the robot, while the social gazing behaviour in Zanatto, Patacchiola, Goslin and Cangelosi (2016)'s experiment only helped human-robot cooperation if the robot was humanoid. Also, Stanton and Stevens (2014) found that robot gaze elicited trust in easier human-robot cooperative tasks, but hindered it in more difficult tasks, suggesting that robot gaze might create uncomfortable pressure in certain situations. By contrast, Ham, Bokhorst, Cuijpers, van der Pol and Cabibihan (2011) found that robot gaze increased its persuasion. Individual differences might partly account for these results, corroborating findings from Mutlu, Forlizzi and Hodgins (2006), who showed that female participants liked robots more when looked at less, while male participants liked robots more when looked at more. Given the contradicting previous results, the current experiment could help examining whether the manipulation of a robot's gaze and attention would affect users' implicit trust attributions of that robot as well.

Voice and attention are two cues that might influence perception of a robot's trustworthiness. However, this influence should also take context and temporal dynamics into account. In the field of HRI, recently Salem et al. (2013) found that robot perception changed drastically after interacting with a robot in two different contexts, and suggested that results based on single robot encounters might not be accurate representations of the perception that people might build over time. Koay, Syrdal, Walters and Dautenhahn (2007) also found that interpersonal distances between a human and a robot decreased over interaction sessions distributed over 5 weeks, and similarly Walters, Syrdal, Dautenhahn et al. (2008) emphasized the importance of long-term studies in HRI, while suggesting that studies focusing on first impressions (such as theirs) are also useful to build a baseline of attributions. Haring et al. (2013) measured trust attributions before and after the interaction with a robot, and found differences in the initial and informed impressions. Haring et al. (2013) also attributed either a trustworthy or untrustworthy behaviour to the robot they employed in their study, but did not explore the effect of this condition fully. Generally, a few of the mentioned authors have raised the issue of the lack of studies involving repeated exposure to a certain agent and to its behaviour in specific tasks (Salem et al., 2013; Walters, Syrdal, Dautenhahn et al., 2008; Koay et al., 2007). This Chapter addresses this lack of empirical evidence on the effect of context and experience on trust in HRI. Previous experiments within this thesis showed that implicit trust attributions to virtual agents are modulated by the agent's voice and behaviour, and that initial impressions changed over time. Here, the voice, attention, and behaviour of a robot are manipulated in order to collect implicit trustworthiness attributions.

Studies on implicit trustworthiness attributions are also scarce in HRI research, with data being mostly collected with questionnaires. However, there are a few studies which collected implicit measures of trust. For example, DeSteno et al. (2012) recorded face-to-face verbal interactions between human participants and a Nexi robot, to identify sequences of non-verbal cues that are indicative of trustworthy behaviour, thus demonstrating that the accuracy of trustworthiness judgments of novel partners is influenced by exposure to nonverbal cues. Haring et al. (2013), among other measures, used proxemics, namely the interpersonal distance between a person and a robot, to measure the person's implicit trust towards the robot. Finally, Hancock et al. (2011) suggested that future research should use implicit data to study Human-Robot trust, and mentioned the problem of individual measures taken after a single interaction, which are not informative of trust development. Thus, scholars in the field of HRI are aware of the lack of results on implicit measures of trust development. The investment game seems like the perfect tool to address this issue.

	Voice	**Attention**	**Behaviour**
1	Natural	Joint	Generous
2	Natural	Joint	Mean
3	Natural	Non-joint	Generous
4	Natural	Non-joint	Mean
5	Synthetic	Joint	Generous
6	Synthetic	Joint	Mean
7	Synthetic	Non-joint	Generous
8	Synthetic	Non-joint	Mean

Table 7: The 8 robot manipulation conditions.

This chapter presents results from an experiment aimed at studying the effect of different anthropomorphic cues — voice and attention — and experience on implicit judgments of a robot's trustworthiness. The robot's voice was either natural or synthetic, the attention was either joint or non-joint and the behaviour was either trustworthy (*generous*) or untrustworthy (*mean*). A summary of the robot's manipulations can be seen in Table 7. The classification of the voice and attention conditions in terms of anthropomorphism is clear, with natural voice and joint attention belonging to the anthropomorphic end of the scale, and synthetic voice and non-joint attention to the non anthropomorphic end of the scale. Manipulating the behaviour condition, instead, is not an assertion that one of the two levels of this condition is more anthropomorphic. Rather, it is meant as a simplistic, categorical representation of two possible behaviours, which, in combination to different voice and attention conditions, might elicit different trusting behaviours from the participants.

Participants in the study were also administered several well known questionnaires after they interacted with the robot, in order to obtain a measure of the perceived anthropomorphism and trustworthiness of the robot after they interacted with it.

8.2 Method

8.2.1 Participants

One hundred twenty individuals (81 female and 39 male; mean age = 23.12 years, SD = 8.62 years) participated in the study. All participants were native British English speakers. Participants were naïve as to the purpose of the investigation and gave informed written consent to participate in the study.

8.2.2 Stimuli

The voices used in this experiment were two female speakers of Standard Southern British English (SSBE). These were chosen among the 4 SSBE voices already used in Experiment 2. There were two blocks of 20 sentences (one for each round of the game), all approximately the same length (mean number of syllables per sentence 16.6, SD 1.08), for each speaker. Participants heard one block of utterances from one robot, and the other block from the other robot, in a counterbalanced order. The two blocks of utterances were equivalent in the *generous* and *mean* conditions. The recorded utterances were amplitude-normalized, and a noise-removal filter was applied.

To obtain the synthetic version of these utterances, the mean f0 value for each utterance was extracted using the software for phonetic analyses Praat (Boersma & Weenink, 2017). The mean f0 value for each speaker was later obtained by averaging the value of all the utterances from the same speaker. Then, the f0 of each utterance was flattened to this mean f0 value. Finally, a comb filter was applied to each flattened utterance using the software Audacity (comb frequency = same value as the flattened f0 of the sound file; comb decay = 0.1; normalization level = 0.990).

Data validation

Prior to starting the experiment, 12 independent native English speakers (who did not subsequently take part in the experiment) were asked to transcribe some of the utterances used in the game, in order to ensure that there were no systematic differences in the intelligibility of the natural and synthetic stimuli. Six of the judges transcribed 10 random natural utterances and the other six transcribed the 10 corresponding resynthesized utterances. The transcription errors that the two sets of judges made were then counted, and a test of proportions was performed to see whether the difference of errors in the natural and synthetic utterances was systematic or due to chance. Overall, the transcribers of the natural speech made 9 errors (4 content words, 5 non content words), and the transcribers of the synthetic speech made 11 errors (3 content words, 8 non content). There were no significantly detectable differences in intelligibility over this sample of speakers ($\chi^2(1) = 0.06$, $p = .806$). Additionally, a sample of the actual participants in the experiment was asked whether they were able to understand the utterances of the robot (in both the synthetic and natural condition); of the 55 randomly chosen participants, only one said they could not understand what the robot was saying, one could understand “very little”, 7 stated they understood most of it, while the remaining 46 said they were able to understand everything. Finally, since only a female voice was used, we asked a random sample of participants to state whether they had associated a gender with the robot. Of the 66 randomly sampled participants, 23 said they thought the robot was female, 17 male, 20 did not associate any gender, and 6 associated a different gender to the two robots they played with; a test of proportions did not yield significant differences between the three groups ($\chi^2(2) = 2.75$, $p = .253$). Thus, it is excluded that participants consistently associated the Nao robot with the same gender, and the current results are generalisable to interactions with a gender-neutral NAO robot.

8.2.3 Procedure

Figure 30: Investment game setup.

Investment game

All the interactions with the robot during the investment game were mediated by a Sandtray touchscreen computer (Baxter, Wood & Belpaeme, 2012), which stood between the participant and the robot (Figure 30). Sandtray is a new platform built to aid the exploration of Human-Robot Interaction. The hardware consists of a 26-inch capacitive touchscreen and associated control server, upon which a series of virtual objects (icons) can be manipulated by tapping (from the human), or simulating dragging (from the robot). The interface linking the robot and the computer was programmed in Python. The display had 11 numbered touch-screen buttons on the side of the participants, which they could press to indicate the amount of money they wanted to invest in each round (from 0 to 10). On the robot's side, there was an orange slider, which was used to give a visual representation of how much money the robot was returning. The starting position of the slider was always at its midpoint, as can be seen in Figure 31. The slider then moved according to the percentage of the robot's returns: it moved to the left of the participant when the robot returned more money than the participant invested, and to the right when it returned less money. All the display was aligned to the participant's visual, to ensure ease of understanding. The display also showed how much money the participant had in the current round, at each time (the "round total" button in the figure). For example, at the start of the round this button showed 10, then, if participants invested 3, it showed 7 and, if the robot returned 6, it showed 13. The "bank" button showed the sum of money that participants had earned in all the previous rounds. The buttons at the top of the screen showed how much money the robot had received and returned. As soon as the participant pressed "start", the first sound file was loaded and played, and participants could not press any button until the file had finished playing. They were prompted to press "start" again to begin each round, at which point the corresponding sound file would be played. This was done to ensure that participants had enough time to look at all the changes in the screen and understand what was going on in the game. After the 20th round, a goodbye message appeared, and participants were directed by the experimenter to the next task.

The robot was standing on one side of the Sandtray computer, facing the participant, who was sitting on the opposite side. The robot was manipulated in terms of voice and attention. In the voice manipulation, the robot had either a natural, pre-recorded female British English voice, or a synthetic voice obtained from the resynthesis of the same British voice, as described earlier (Section 8.2.2). In the attention condition, the robot's arm and head movements were manipulated. In the joint attention condition, the robot located the face of the participant and its head followed the participant's movements, giving an impression of looking at the participant. This was achieved with the built-in Nao face detection and tracking algorithms. When it was the robot's turn in the game, it lowered its head, "looking" at the Sandtray screen, and performed a sweeping arm movement, following the slider that, on the screen, indicated how much money the robot was returning. In the joint attention condition, the robot's right

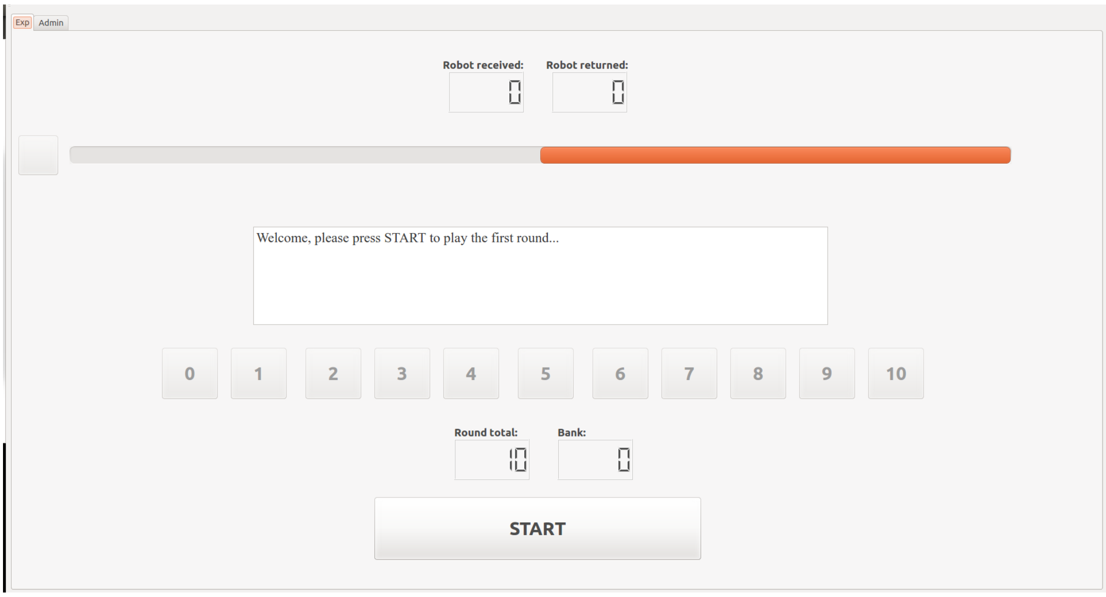


Figure 31: Sandtray display at the beginning of the investment game.

arm moved to its right, following the slider, when it was returning more money, while its left arm moved to the left, following the slider, when it was returning less money. Whenever the arm moved, the robot also pointed at the slider with a finger, giving the impression that it was controlling the slider in this way. In the non-joint attention condition, instead, the robot was standing still with its head lowered, and never looked at the participant, thus giving the impression that it was always looking at the screen.

Before the first game started, the experimenter ran 3 practice rounds with the participant, in order to familiarise them with the interface. Then, the experiment followed the same procedure as the previous four experiments. Each participant engaged in two games, one for each voice, attention and behaviour condition (*generous*/*mean*), with a different block of 20 utterances heard for each version of the game. The experiment was counterbalanced in a 2 (attention: joint or non-joint) by 2 (speaker: 1 or 2) within subject design and a 2 (behaviour: *generous* or *mean*) by 2 (voice: natural or synthetic) between subject design. The game rounds proceeded as in the previous experiments. In order to increase the intrinsic motivation of participants, they were told that they would win actual money in proportion to their earnings in the game. For this reason, they played with “Experimental Currency

Units" (ECU), receiving 10 at the beginning of each round. The total amount that they earned was converted in British Pounds, at a rate of 30 ECU = £0.10, and this was paid to the participants at the end of the game. In the previous experiments, return ranges were 120-240% in the *generous* condition and 0%-120% in the *mean* condition. As previously discussed (Section 3.3), this might have induced participants to trust more in the *mean* condition. In order to examine whether participants would be less trusting of a completely untrustworthy game partner, this range was reduced to 150%-240% (*generous* condition) and 0%-90% (*mean* condition).

At the end of the first game, participants were told that they would now play a second game with another robot, and were asked to complete a set of questionnaires (described below), while the experimenter brought in the second robot. In reality, while the participant answered the questionnaires, the experimenter unplugged the Nao robot, went to an adjacent room, and changed the T-shirt that the robot was wearing, before returning to the experiment room. The T-shirt was either red or blue, and it was meant to give the impression that the robot was different from the previous one.

Questionnaires

Four questionnaires were used as secondary measures to the main game task. Three short scales measured likeability, trust, and credibility. The likeability questionnaire was based on Reysen (2005), while the trust scale was an adaptation of the Receptivity/Trust subscale of the Relational Communication Questionnaire and of the selection of trust items in the International Personality Item Pool (Goldberg et al., 2006, IPIP,). The credibility scale was based on McCroskey and Young (1981)'s Source Credibility Scale. In addition, Bartneck, Kulić, Croft and Zoghbi (2008)'s questionnaire was used to measure a range of HRI factors (anthropomorphism, animacy, likeability, perceived intelligence and perceived safety). After the first game ended, the participants were asked to fill in the questionnaires, while the experimenter took the robot to another room. While the participants were told that the experimenter would bring in a different robot, in reality the experimenter changed the t-shirt that the robot was wearing. This trick was meant to make the participants believe that the second robot he/she would be playing with was essentially a new agent, and as such the participant's strategy would reset. Participants filled in a second set of questionnaires at the end of the second game. Finally, participants completed a short background questionnaire where they were asked their age, gender, city of origin, what accent they spoke, and what accent they would like a robot to have. They were then debriefed, paid the show-up fee plus what they had earned in the game, and left.

8.3 Results

8.3.1 Investment game results

A linear mixed-effects model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC (Aho et al., 2014, Akaike Information Criterion,), with investment as dependent variable, speaker, behaviour (*generous*/*mean*), voice (natural/synthetic), attention (joint/non-joint) and game turn (the 20 rounds of the game) as independent variables, and participant id and sentence block as random factors.

Main effects

As expected, there was a main effect of behaviour ($\chi^2(1) = 90.17$, $p < .001$), with higher investments in the *generous* condition (mean = 8.13 ECU) than in the *mean* condition (mean = 4.5 ECU), as can be seen from Figure 32. There was also a main effect of game turn ($\chi^2(1) = 26.58$, $p < .001$), with overall higher investments in the second half of the game. There was also a main effect of voice ($\chi^2(1) = 4.47$, $p < .05$), with higher investments with the synthetic voice. There was also a main effect of speaker ($\chi^2(2) = 7.37$, $p < .05$), with higher overall investments with speaker number 2. There was no effect of attention however ($\chi^2(1) = 0.41$, $p = .52$).

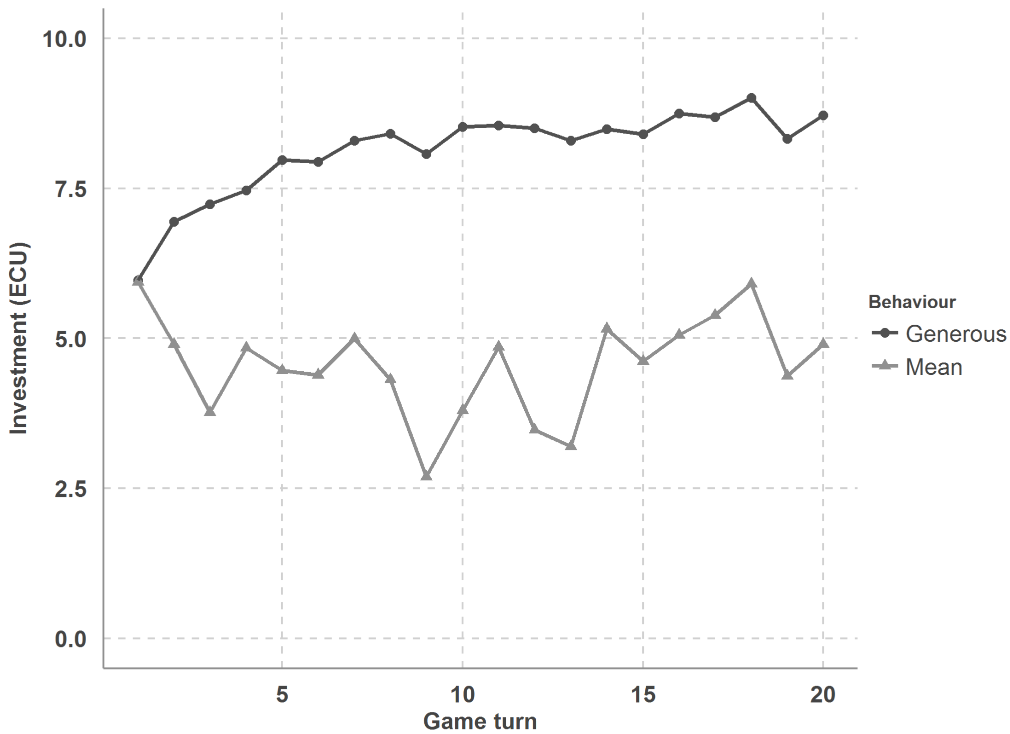


Figure 32: Average investments to the *generous* and *mean* conditions.

Interactions

There was an interaction between behaviour and game turn ($\chi^2(1) = 18.26$, $p < .001$): as can be seen from Figure 32, investments increase over time in the *generous* condition in an almost linear fashion, but fluctuate heavily in the *mean* condition. There was also an interaction between behaviour, game turn and voice ($\chi^2(2) = 8$, $p < .02$). A post-hoc mixed-effects model was fitted to the *generous* and *mean* data separately to analyse this interaction. Investment was the dependent variable, game turn and voice were predictors, and participant and sentence block were random factors. In the *generous* condition, there was a main effect of game turn ($\chi^2(1) = 256.5$, $p < .001$) and a marginally significant effect of voice ($\chi^2(1) = 3.32$, $p = .068$). As can be seen from the top panel in Figure 33, investments increase over time, and are higher in the synthetic voice condition. In the *mean* condition, instead, there is no effect of game turn, and no effect of voice, but there is a significant interaction between game turn and voice ($\chi^2(2) = 16.33$, $p < .001$). As can be seen from the bottom panel of Figure 33, investments decrease in the first half of the game independently from the voice, while they increase in the second half of the game, and there are higher investments to the natural voice.

There was also an interaction between behaviour and speaker ($\chi^2(1) = 6.38$, $p < .02$), with higher investments in speaker 2, but only in the *mean* condition, and between voice and speaker ($\chi^2(1) = 10.43$, $p < .002$), with higher investments in speaker 2, but only with a synthetic voice. There was also a marginally significant interaction between game turn and attention ($\chi^2(1) = 3.66$, $p = .056$, significant in terms of AIC): participants invested more money in the second half of the game in both the attention conditions, although the slope was steeper for the joint attention condition, as shown in Figure 34. There was also an interaction between speaker and attention ($\chi^2(1) = 10.98$, $p < .001$), with higher investments to speaker 2, but only in the non joint attention condition. Finally, there was an interaction between behaviour, speaker and attention ($\chi^2(2) = 7.92$, $p < .02$). A post-hoc mixed-effects model was fitted to the *generous* and *mean* data separately to analyse this interaction. Investment was the dependent variable, speaker and attention

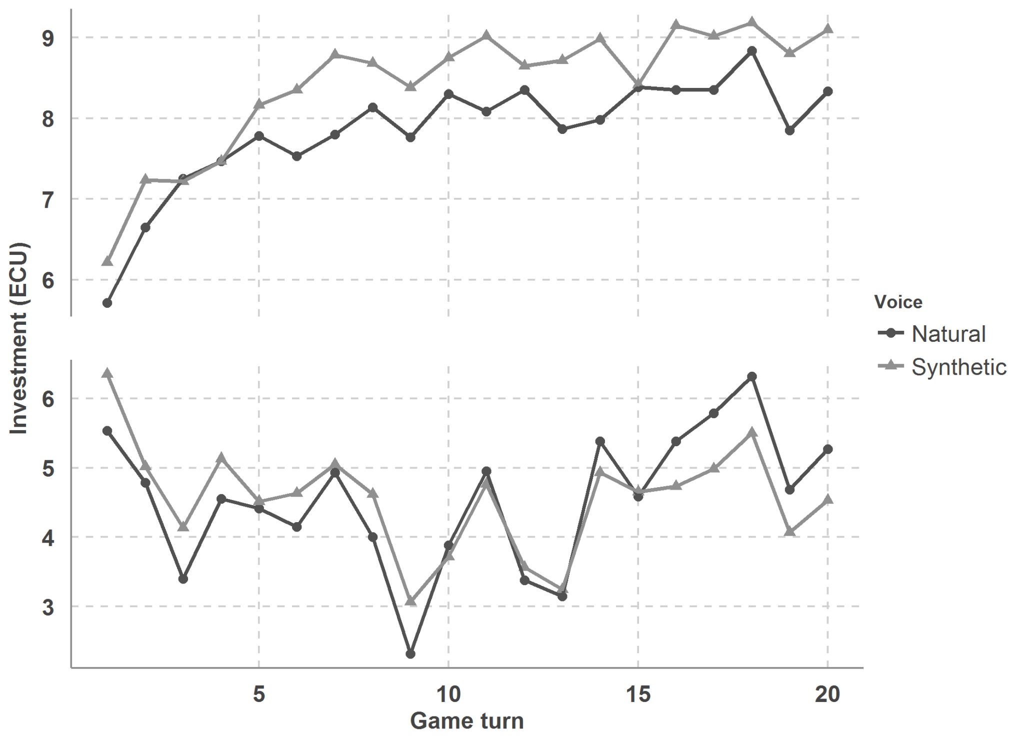


Figure 33: Average investments in the two different voice conditions and in the *generous* (top) and *mean* (bottom) behaviour conditions.

were predictors, and participant and sentence block were random factors. In the *generous* condition, there were no main effects of speaker or attention, but there was a significant interaction between the two variables ($\chi^2(2) = 14.4$, $p = .002$): participants invested more money to speaker 1 than to speaker 2 in the joint attention condition (average = 8.3 ECU and 7.96 ECU, respectively), while they invested more in speaker 2 than speaker 1 in the non-joint attention condition (average = 8.32 ECU and 7.91 ECU, respectively). In the *mean* condition, instead, there was a main effect of speaker ($\chi^2(1) = 8.38$, $p = .004$), with higher overall investments to speaker 2 than speaker 1 (average = 4.68 ECU and 4.41 ECU, respectively). There was no effect of attention, and no interaction between speaker and attention.

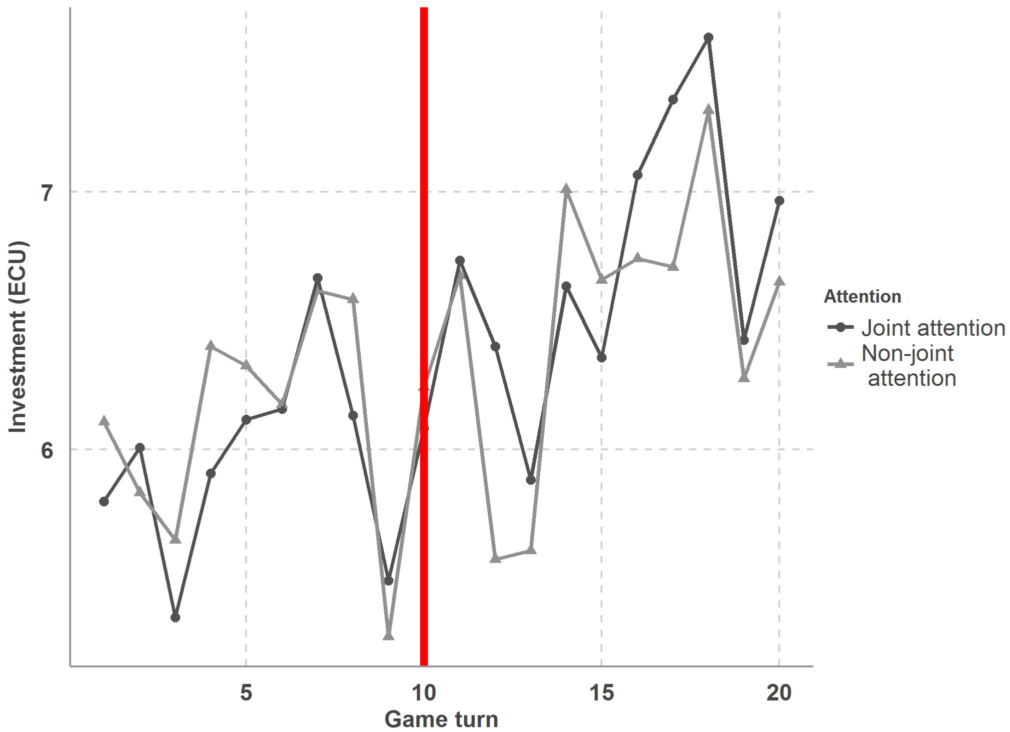


Figure 34: Average investments in the two attention conditions over time. The red vertical line indicates the mid-point in the game.

	Behaviour	Attention	Voice	Speaker
Likeability	$\chi^2(1) = 32.67$, $p < .001$	$\chi^2(1) = 14.14$, $p < .001$	n.s.	n.s.

	Behaviour	Attention	Voice	Speaker
Trust	$\chi^2(1) = 43.65$, $p < .001$	$\chi^2(1) = 27.27$, $p < .001$	n.s.	n.s.
Credibility	$\chi^2(1) = 43.65$, $p < .001$	$\chi^2(1) = 15.43$, $p < .001$	n.s.	$\chi^2(1) = 4.53$, $p = .033$
		Godspeed Questionnaires		
Anthropomorphism	n.s.	$\chi^2(1) = 30.21$, $p < .001$	n.s.	n.s.
Animacy	n.s.	$\chi^2(1) = 66.61$, $p < .001$	n.s.	n.s.
Likeability	$\chi^2(1) = 71.91$, $p < .001$	$\chi^2(1) = 16.85$, $p < .001$	n.s.	n.s.
Intelligence	$\chi^2(1) = 18.13$, $p < .001$	$\chi^2(1) = 6.17$, $p = .012$	n.s.	n.s.
Safety	n.s.	$\chi^2(1) = 8.76$, $p = .003$	n.s.	n.s.

Table 8: Main effects of questionnaires.

8.3.2 Questionnaires Results

For all the scales, a linear mixed-effects model was fitted to the data, with behaviour (*generous*/*mean*), voice (natural/synthetic), speaker and attention (joint/non-joint) as independent variables, and participant id and sentence block as random factors. Post-hoc comparisons, where needed, were assessed using t-tests.

Main effects

As shown in Table 8, behaviour had an effect for the likeability, trust, credibility and intelligence scales, with higher ratings for the *generous* robot. Attention affected all the scales, showing higher scores for the joint attention condition. On the contrary, voice had no effect, while speaker had an effect only in the credibility scale, with higher ratings for speaker 1.

Interactions

All the two-way interactions that were found are reported in Figure 35. Both likeability scales showed a significant two-way interaction between behaviour and voice ($\chi^2(1) = 8.83$, $p = .002$; $\chi^2(1) = 7.36$, $p = .006$), with higher likeability ratings for the synthetic voice in the *generous* behaviour condition ($p = .02$), and for the natural voice in the *mean* behaviour condition ($p = .027$). For trust and safety a two-way interaction between behaviour and attention emerged. For the former ($\chi^2 = 4.01$, $p = .004$), joint attention was rated as more trustworthy for both *generous* ($p < .001$) and *mean* behaviours ($p = .021$), while for the latter ($\chi^2(1) = 8.84$, $p = .002$) the joint attention condition was perceived as safer than the non-joint one in the *generous* behaviour condition ($p < .001$), but no differences for attention were found ($p = .995$) in the *mean* behaviour condition. Moreover, in the safety rating there was a significant interaction between voice and speaker ($\chi^2(1) = 6.30$, $p = .012$). Post-hoc comparisons showed that the synthetic voice was rated as safer than the natural one when the voice belonged to speaker 1 ($p = .01$). For speaker 2, no significant effect of voice emerged ($p = .43$). There was a significant interaction between behaviour and speaker in the anthropomorphism ($\chi^2(1) = 6.82$, $p = .008$) and animacy ($\chi^2(1) = 7.18$, $p = .007$) scales, with speaker 1 being rated as more anthropomorphic in the *mean* behaviour condition ($p = .012$ for anthropomorphism, $p = .038$ for animacy), but not in the *generous* ($p = .49$ for anthropomorphism, $p = .16$ for animacy).

Finally, there were also some significant three-way interactions. In the anthropomorphism scale there was an interaction between voice, speaker and attention ($\chi^2(1) = 4.26$, $p = .038$). Specifically, voice had an effect on the second speaker in the joint attention condition, where the natural voice was rated as more anthropomorphic than the synthetic one ($p = .012$). All the other comparisons were not significant ($p > .05$). In the intelligence scale there was an interaction between behaviour, speaker and attention ($\chi^2(1) = 4.27$, $p = .038$): speaker had an effect in the *mean* and non-joint attention condition ($p = .006$), where intelligence was higher with speaker 1. All the other comparisons were not significant ($p > .05$). Finally, there was also a significant interaction between behaviour, voice and attention in the Safety scale

($\chi^2(1) = 9.86$, $p = .001$). Specifically, voice had an effect in the *generous* and non-joint attention condition ($p = .015$), with lower safety ratings with the natural voice. All the other comparisons were not significant ($p > .05$).

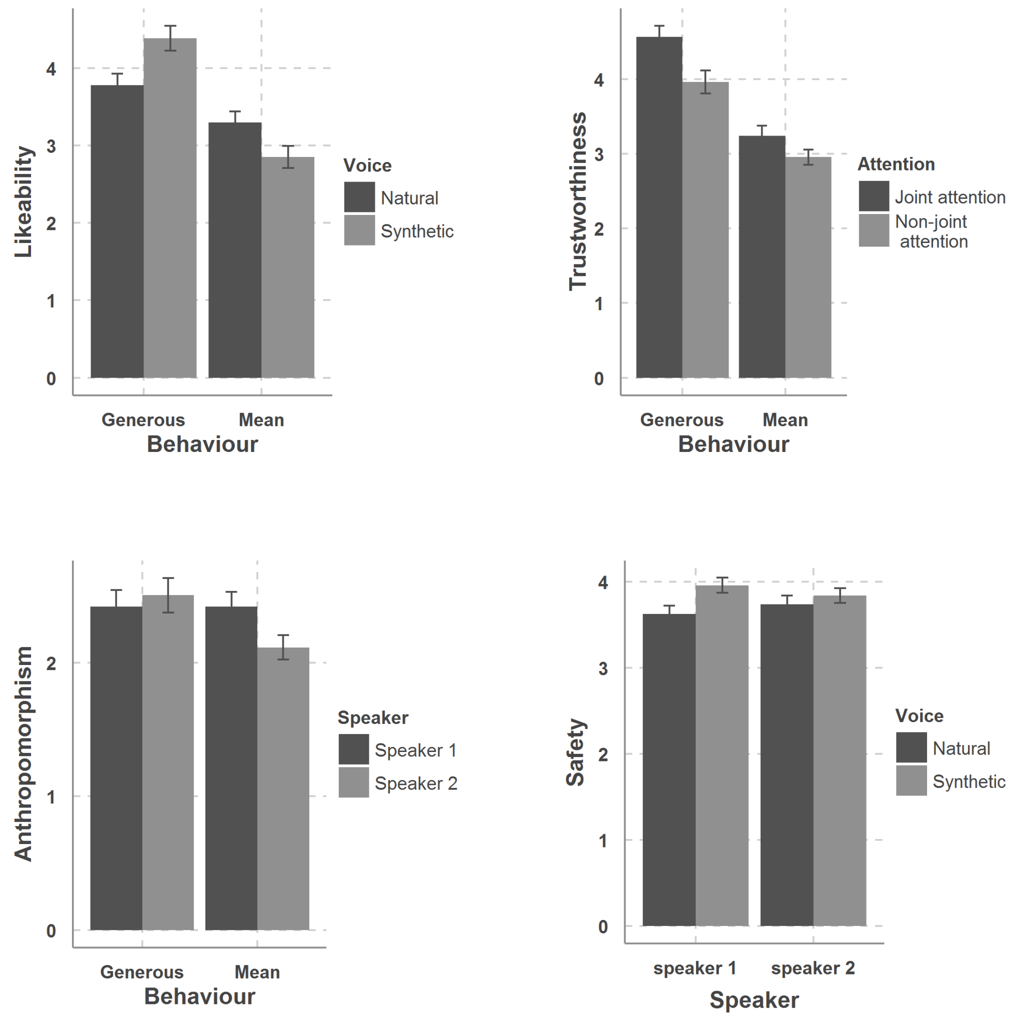


Figure 35: Two-ways interactions for the likeability, trust, anthropomorphism and safety scales.

8.4 Discussion

The data reported here show that people's implicit trust attributions to a humanoid robot change based on different anthropomorphic cues, and over time, based on experience. Some of these anthropomorphic cues also influenced results in several perception questionnaires that participants filled after interacting with the robot.

In the game, participants' investments were modulated by the behaviour and voice of the robot, and by the particular speaker to whom the original voice belonged. Participants invested consistently more money with a *generous* robot, with a robot with a synthetic voice, and with a robot whose voice was taken from speaker 2. Higher investments with a *generous* game partner were expected, as the goal of the participants was ultimately to earn more money, and as such they learned very quickly whether the partner was being trustworthy or not. Regarding auditory cues, it seems that participants preferred one particular speaker over the other. 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 (see Section 2.2.3; also Gobl & Ní Chasaide, 2003; Trouvain et al., 2006). This also suggests that people's preference for certain individual voices might apply when these voices are embodied in a robot. Thus, robot manufacturers should not only concentrate on higher-level features of a voice, such as its naturalness and accent, but also on idiolectal characteristics, since they seem to contribute to trusting behaviours as well. Also, participants overall trusted voice congruent robots more than incongruent ones, i.e. they preferred the humanoid robot Nao to have a synthetic voice rather than a natural one. The other anthropomorphic cue that was manipulated, attention, had no effect on the game results, however.

Regarding the congruency of the attention condition, the interaction that was found between game turn and attention supports results that participants prefer interacting with a robot exhibiting human-like characteristics in terms of hand pointing and gaze behaviour (Admoni et al., 2011; Staudte & Crocker, 2011), albeit this effect manifests itself after some time. On the one hand, this result confirms previous calls for studies on the long-term effect of human-robot interactions (Hancock et al., 2011; Salem et al., 2013; Walters, Syrdal, Dautenhahn et al., 2008; Koay et al., 2007). On the other, it suggests that vocal and visual characteristics might be processed differently by participants, and thus the congruency of each of them should be considered separately. Also, the interaction that was found between speaker and attention shows that preference towards one particular speaker only emerged in the non-joint attention condition, thus suggesting that kinaesthetic information (when present) might override fine speech information.

The results from the investment game also support previous studies suggesting that robots are treated as conscious agents (Nass & Lee, 2001; Nass et al., 1994; Walters, Syrdal, Dautenhahn et al., 2008). In fact, if people had felt less connection with the robot as an agent, they might have invested lower amounts of money, especially in the *mean* condition, indicative of an unwillingness to cooperate. Instead, the patterns of investments in the two behaviour conditions are similar to the patterns of investments in the previous experiments, which used disembodied virtual agents with only natural voices as a cue to their agency. In these studies, participants' investments in the *generous* condition were increasing almost linearly over time, while investments in the *mean* condition fluctuated heavily, similarly to the present study (Figure 32). This fluctuation of investments could be a strategy from the human participant to get the robot to return more money, by investing more money in the first place; this is a strategy that has sometimes been observed in human-human trust games (Buchan et al., 2008; Macy & Skvoretz, 1998). This fluctuation is further discussed in Section 9.4, with reference to all the experiments presented here.

Intriguingly, higher investments in the *mean* condition were directed towards the robot with a natural voice, possibly signalling that, while participants felt there was no hope of convincing the synthetic-voiced robot to return more money, they felt more hopeful that this could be accomplished if the robot had at least a human-like trait, its voice. This refers back to Theory of Mind: the robot's voice might have been the difference between perceived agency and lack thereof; if participants believed that the robot did not have any intentions, this could help explain the lower investments to the synthetic-voiced robot in the *mean* condition.

Behaviour, voice and speaker, the characteristics which most affected participants' investment decisions in the game, had little effect in the post-game questionnaires. Thus, it is possible that, while participants are unconsciously aware of all the robot's characteristics, some of these might be more important during the actual interaction, whereas others might be more predominantly recalled after the interaction has taken place. The type of interaction might be a conditioning factor as well, since in this case the interaction was verbally mediated, even though the participants were not meant to talk back to the robot. Participants' investments also increased more rapidly in the second half of the game in the joint attention condition, thus suggesting that joint attention started having an effect in the investment decisions only after participants had been exposed to it for some time. While joint attention had no main effect in the game, it was consistently linked to participants' explicit judgments in the questionnaires. In particular, it was a significant predictor of perceived likeability, trustworthiness, credibility, anthropomorphism, animacy, intelligence and safety of the robot. Some of the interactions that were found in the game were also replicated in the questionnaires. For example, participants rated a synthetic voice higher in likeability if they had just heard it in conjunction with the *generous* behaviour, and a natural voice if it was paired with the *mean* behaviour. The questionnaire results are also in line with previous literature: Eyssel, Kuchenbrandt, Hegel and de Ruiter (2012), for example, also found that the type of robot voice they used (natural or synthetic) had no effect on subsequent anthropomorphic ratings of the robot. It is possible that participants turned their attention to the movement and gazing behaviour of the robot only when they were asked to reflect on it; in the case of the questionnaires, they might have been socially biased to care about these features more than the voice. The reason behind this different processing of auditory and visual cues during and after an interaction remains to be explored.

The design and results of the present experiment allow to at least speculate on the reason behind such a differential effect of the two anthropomorphic cues that were manipulated.

First of all, it is possible that participants attended to the speech cues more than to the visual cues. Participants' auditory channel was free to attend to the robot's speech all the time, without distractions, while their visual attention likely switched between the robot and the shared screen. Thus, while participants had a constant exposure to the robot's voice, details of the robot's movements might have gone unnoticed from time to time. Also, even though the utterances spoken by the robot did not have any connection with what was happening during the game, it is possible that participants attributed some importance to them, thus allocating more working memory to the processing of this — effectively useless — information.

It is also possible that, while vocal characteristics have an immediate effect on the users' implicit perception of the robot, visual characteristics might take some time to get used to and to be noticed. This would explain the more rapid increase of investments in the joint attention condition only in the second half of the game. This would also fit with the hypothesis that participants might have taken some time to familiarise themselves with the game interface. At the beginning of the game, participants might have been more concentrated on the game and its interface, rather than the robot, and only after some time they paid more attention to the robot's movements.

Moreover, while the voice condition was clearly split into a natural and synthetic category, it might be that the joint attention condition was not perceived as comparatively human-like as the voice. If this was the case, the distinction that had been allocated *a priori* to the joint and non-joint attention conditions would be comparatively different to the distinction between the natural and synthetic voice. Although the robot's arm movement and face tracking were clearly distinguished from the non-joint attention condition, where the robot was not moving at all, it is possible that such movements were still perceived as mechanical. The robot's gaze might also have caused discomfort, as in Stanton and Stevens (2014), thus contributing to the elimination of differences between the still and moving robot. It is also possible that this difference in feature effect might be due to a confound between type of measure (implicit and explicit) and type of task (investment game and various questionnaires). Specifically, the investment game arguably measures implicit trusting behaviour in a specific setting, which has been linked to some, but not all, facets of trust (Ben-Ner & Halldorsson, 2010). On the other hand, the questionnaires used measure explicit ratings on several different scales, only one of which is specifically about trust.

Finally, participants' explicit understanding of the word "trust" (used in the questionnaire) might capture different shades of meaning than the economic trusting behaviour predominantly used in the game, and these different shades might be linked to different cues as well.

The present experiment has several implications. First of all, it seems that people implicitly attend to vocal cues in human-robot interactions. In general, participants in this experiment invest more money with a synthetic-voiced robot, within game modulations: the preference towards a synthetic/congruent voice only lasts until things start deteriorating and participants realise that the robot is behaving untrustworthily. At this point, implicit trust in the natural voice condition becomes higher, perhaps in an appeal to human empathy. Voice does not affect explicit perceived trustworthiness of the robot, although it does affect other perceptual evaluations. For instance, perceived likeability increases for the synthetic voice when the robot is *generous*, and conversely the natural voice is preferred when the robot becomes *mean*. However, implicit game choices are not just reflections of trust, but also attempts to signal behaviour and trust in another. Questionnaires are merely reflective, and cannot be used to try to modify the behaviour of a counterpart. Thus, the robot's voice had a strong influence on actual trusting behaviours.

Secondly, preference towards a robot exhibiting joint attention behaviours emerges with time: in the second half of the game, and in the post-game questionnaires. It is possible that it takes time to get used to human-like movements, which end up being preferred to the completely non-human motionlessness.

Also, it appears that participants pay greater attention to speech cues when cooperative solutions are not being met. When the robot is behaving favourably (i.e. it is being *generous* in the game), participants seem not to care about its individual features such as the voice. It is when things start deteriorating, i.e. when the robot is *mean*, that we start paying attention and attending to anthropomorphic cues. This is when individual preferences to speech features start as well (e.g. the preference of one of the two speakers over the other). Just like in the game, the anthropomorphism and animacy scales show a preference for one of the speakers in the *mean* condition. All taken together, this evidence suggests that participants need to rely on or appeal to human-like features when things are not going the desired way.

Finally, the current results suggest that implicit and explicit measures might both contribute to understanding how trust is formed, although they might describe different aspects of it. For example, there was an opposite effect of speaker in the game and in the credibility questionnaire, no effect of attention in the game, and no effect of voice in the perceptual questionnaires. While the game and the questionnaire data arguably measure different aspects, they demonstrate the complexity of the multifaceted concept of trust and the necessity to use both instruments to better outline its structure.

8.5 Conclusion

This chapter reported implicit and explicit trust attributions to a humanoid robot whose voice and joint attention were assessed during the experiment. These features were manipulated to be more or less anthropomorphic, and the robot's behaviour was manipulated to be more or less trustworthy during an economic game. The underlying idea was to study the effect of robot anthropomorphic design on human perception and trust, as this is not yet widely understood. While some design features such as physical appearance have been studied in depth (Scassellati, Admoni & Matarić, 2012; Powers & Kiesler, 2006; Hinds, Roberts & Jones, 2004; Zhang, Zhu, Lee & Kaber, 2008; Walters, Syrdal, Dautenhahn et al., 2008), others, for example voice and movement, have received less attention in the past (Lee et al., 2006; Walters, Syrdal, Koay, Dautenhahn & Te Boekhorst, 2008; Staudte & Crocker, 2011).

The present findings suggest that, while the feature that was most prominently linked to the robot's "being" — its voice — had a strong effect on participants' strategies all throughout the game, the feature that was mostly linked to the robot's "state" — its movement — influenced trusting behaviours mostly in the second half of the game. Conversely, the questionnaire results indicate that the robot's movement and attention had the biggest effect on perceptual evaluation of the robot's likeability, trustworthiness, credibility, anthropomorphism, animacy, intelligence and safety, while robot's voice and behaviour had a limited effect. Taken together, these results suggest that people attend to certain cues more than others in the process of deciding whether to trust a robot or not, while other cues might be more predominantly recalled when reflecting on a past interaction.

This research shows that general design features such as the robot's voice might be context and behaviour-dependent, and opens new interesting research questions. For example, how would a robot's synthetic accent interact with context and behaviour? Also, should vocal manipulations such as pitch and articulation rate be congruent with the physical appearance of the robot? Future research should address this topic more specifically.

Some semi-anecdotal evidence collected in the 5 experiments presented so far should at least partially answer the first open issue. The final question that participants answered in the post-game background questionnaire in all the experiments was: "What accent would you like a robot to have?". Figure 36 shows the standardised answers of all 503 participants who answered the question. As the figure shows, the majority of respondents answered with "SSBE", followed by "Neutral" accent (whatever this might mean), followed by "Irish". Very few people answered that they would prefer the robot to have a robot-like voice, or an accent like their own. However, many manufacturers of social robots (Nao *in primis*) only give them a robot-like default voice. While this data by no means represents a scientific claim that robots should have an SSBE or Irish accent, it at least shows that robot manufacturers should pay some attention to the voice they give their robots, since potential users certainly do.

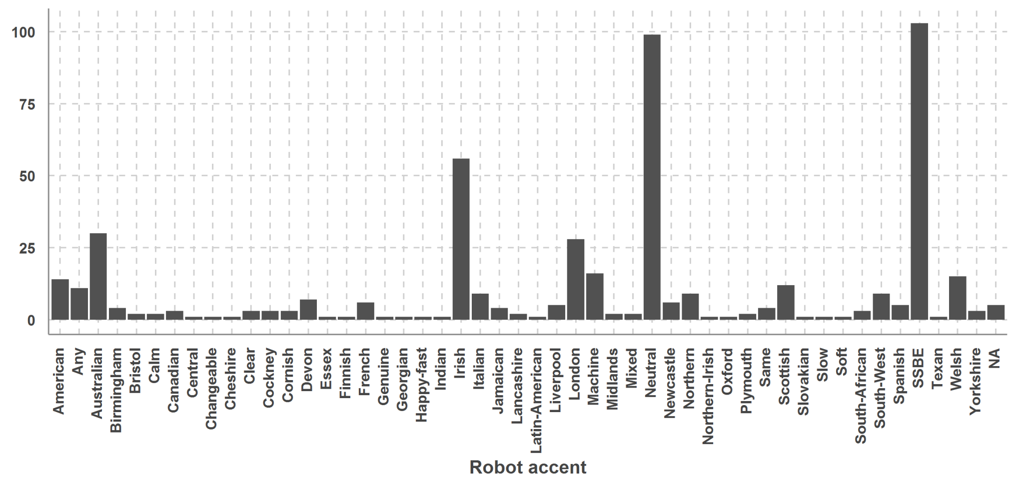


Figure 36: Preference for a robot's accent of 503 participants.

9 General discussion

In this chapter, the results from the 5 Experiments presented in this thesis are summarised and discussed, specifically in terms of the implications they have for speech research and behavioural economics. Potential practical applications of these results are also outlined.

9.1 Summary of findings

This thesis contains 5 experiments aimed at examining the influence of various voice characteristics – including accent, prosody, emotional expression and naturalness – on trusting behaviours towards virtual players or robots.

In Experiment 1, trusting behaviours were dictated by a combination of accent and speaker's trustworthiness, so that people trusted an SSBE-accented player more when he was *generous*, and a Liverpool-accented player more when he was *mean*. In Experiment 2, people trusted SSBE- and Birmingham-accented player the most when they were *generous*, but Birmingham-accented players were trusted the least when they were *mean*. Also, high pitch and slow articulation rate increased trust. In Experiment 3, people trusted speakers with a manipulated high pitch and slow rate more, and trusted female speakers more than male speakers. In Experiment 4, people trusted players who spoke with a smiling voice more. They also trusted SSBE-accented players more when they were *generous*, and Birmingham-accented players more when they were *mean*. In Experiment 5, people overall trusted a Nao robot with a synthetic voice more than one with a natural voice, but they trusted a robot with a natural voice more when it was *mean*.

From these experiments, it is possible to conclude that accent, prosody, emotional expression and naturalness all influence implicit trustworthiness judgments. Furthermore, these characteristics interact with the speaker's actual trustworthiness in the participants' trust development process.

9.2 Voices

The particular vocal features examined in these experiments were accents, prosodic cues, emotional expression and naturalness. As previously discussed, accents are pronunciation differences within a language that distinguish groups of speakers. Prosodic cues, such as pitch or articulation rate, are individual variation in the suprasegmental features of speech. Emotional expression, such as smiling voice, is the conveying of an emotion, such as amusement, in the speech channel. Finally, naturalness is the degree to which a voice sounds "human", as opposed to "synthetic" or "machine-like". All these features influenced trusting behaviours.

From the experiments presented here, it is evident that voices influence implicit trustworthiness judgments. This influence clearly overcame any experiential learning that was due to the trustworthiness or untrustworthiness of the virtual players in the games. The voice, in its different manipulations across experiments, was the only feature that distinguished one trustee from another, and there were no visual or other cues that could further identify them as distinct agents. Indeed, if participants had not been affected by the speakers' voices when making the trusting decisions, there would be no differences in the investments to the different speakers within the two trustee's behaviour conditions.

Overall, the accent of the trustee had consistent effects on people's trusting behaviours. Participants implicitly trusted some accents more than others, even though they did not systematically recognise their origin in post-game questionnaires. Specifically, SSBE speakers were overall trusted more than other regional speakers, both male and female — although their trustworthiness was also strongly influenced by their behaviour, as discussed below.

This suggests that listeners find standard-accented speakers implicitly more trustworthy. Also, this trusting behaviour seems to be independent of the trustors' own accent, thus suggesting that accent familiarity or loyalty effects previously found in sociolinguistics studies (e.g. Giles, 1971) do not influence implicit trust.

In addition to the influence of accent, trusting behaviour is also influenced by pitch and articulation rate, both in natural (Experiment 2, Chapter 5) and in manipulated speech (Experiment 3, Chapter 6). High pitch and slow articulation rate, both for male and female speakers, increased participants' trust. These effects have been discussed in terms of biological primitives, such as the "Size/Frequency" code and the "Effort Code", as well as in terms of persuasion theories.

Emotional expression, in the form of smiling voice, also increased trusting behaviour. In line with "halo effect" theories, this suggests that expressing a positive emotion leads to being attributed other positive traits, such as trustworthiness. As a result of this research, these theories are extended to implicit trustworthiness attributions.

In the case of voices embodied in a Nao robot, people overall trusted a natural voice more than a synthetic one, and also trusted one of the speakers more than another. This suggests that not only the type of voice (whether natural or synthetic), but also individual speaker characteristics should be considered when designing a robot's voice.

Many of the vocal characteristics that were found to influence trust were constant over time. For example, participants in Experiment 1 invested more in the SSBE-accented speaker all throughout the game when the virtual player was trustworthy and more in the Liverpool-accented speaker all throughout the game when the virtual player was untrustworthy. Relative trustworthiness of accents was constant also in Experiment 2, where participants invested more to the SSBE- and Birmingham-accented virtual players when they were trustworthy, and less to the Birmingham-accented virtual players when they were untrustworthy. The lack of interactions between these accents and game turn suggests that preference for certain speakers does not fade away over time.

In terms of prosodic cues, the effect of high pitch on investments was constant over time in Experiment 3, while in Experiment 2 it was stronger earlier in the game, and then diminished over time. This could be due to the fact that speakers in Experiment 2 also had different accents, so participants had more auditory information to process, and might have reacted to lower-level auditory cues such as pitch only after some time. Instead, in Experiment 3 the accent and idiolectal characteristics of the speakers were constant, so participants were able to discriminate exclusively between prosodic cues. The effect of articulation rate, on the other hand, did not interact with game turn in Experiment 2, but it did in Experiment 3, where participants invested higher amounts of money in the slow rate condition later in the game. While the effect of rate in Experiment 2 was rather marginal, it was instead strong in Experiment 3. This discrepancy between the temporally invariant effect of pitch, and the temporally variant effect of articulation rate, suggests that these two prosodic cues might be processed separately. While pitch seems to have an immediate and lasting effect, slow speakers might only be appreciated after some time. Smiling also increased investments throughout the game, and the effect of natural voice on the Nao robot in Experiment 5 also did not diminish with time.

9.3 A "congruency effect" of trusting behaviour to voices

The pre-programmed behaviours in the investment game influenced participants' trusting behaviours towards different speakers. Specifically, in Experiment 1 participants trusted the SSBE-accented virtual player more in the *generous* condition, and the Liverpool-accented virtual player more in the *mean* condition (see Figures 4 and 5). In Experiment 2, participants trusted the SSBE- and Birmingham-accented virtual players the most in the *generous* condition, and the Birmingham-accented virtual players the least in the *mean* condition (Figure 9). In Experiment 4, participants trusted the SSBE-accented virtual player more in the *generous* condition, and the Birmingham-accented virtual player more in the *mean* condition (Figure 22). Also, in Experiment 5, participants trusted a robot with a synthetic voice more in the *generous* condition, and a robot with a natural voice more in the *mean* condition, especially in the second half of the game (Figure 33). These results suggest that participants might form an impression of a speaker’s trustworthiness upon hearing them for the first time – which in the investment game is simulated by the first round of the game. The actual trustworthiness of the speaker, in the form of the simulated *generous* and *mean* behaviours, then contributes to the further shaping of this impression, and this impression is different based on whether the perceived and actual trustworthiness agree or not. For example, participants in Experiment 1 might have attributed trustworthiness to the SSBE speaker when they heard him for the first time. In the *generous* behaviour, this impression of trustworthiness is confirmed, so their trust in this speaker is maintained and reinforced over time. In the *mean* condition, instead, the impression of trustworthiness is disproved, so participants “punish” this incongruity by investing less money in this speaker. On the other hand, participants might have initially attributed a lower trustworthiness to the Liverpool speaker, upon hearing him for the first time. In the *generous* condition, the actual behaviour of the speaker is at odds with first impressions, so participants might still be suspicious of this incongruent behaviour, resulting in lower investments. In the *mean* condition, instead, participants’ implicit first impressions are confirmed, so their reaction to untrustworthy behaviour is milder, because the Liverpool-accented virtual player is behaving congruently with participants’ first impressions.

The argument that a discrepancy between first impressions and actual trustworthiness influences trusting behaviour also applies to the other experiments where an interaction between voice and behaviour was found. In Experiment 2, the Birmingham-accented virtual players received the highest investments in the *generous* condition, and the lowest in the *mean* condition. Thus, participants might have formed a first impression of trustworthiness upon hearing them for the first time, which was reinforced in the *generous* condition and “punished” in the *mean* condition, when their behaviour was perceived to be incongruent. This interaction shows that this “congruency effect” seems to be unrelated to standard-accentedness, but rather it seems that it is applied to voices as a whole. Accent, prosody, idiolect, emotional expression, all contribute to identifying a unique voice. Listeners might then form an impression about a voice trustworthiness based on a combination of different factors, and this first impression is then reinforced or disproved with experience of actual trustworthy or untrustworthy behaviour.

A similar pattern of investments was also found in Experiment 4, where participants trusted the SSBE-accented virtual player more in the *generous* condition, and the Birmingham-accented virtual player more in the *mean* condition. While this result seems to contradict the previously discussed results of Experiment 2, it does not: the SSBE speakers whose voices were used in Experiments 1, 2 and 4 were different, and so were the Birmingham speakers whose voices were used in Experiments 2 and 4. Thus, participants do not trust all speakers of the same accent in the same way, but rather the speaker's accent, together with other vocal characteristics, contributes to their perceived trustworthiness. This result also confirms that participants change their trusting behaviour due to a congruency or discrepancy between their first impression about a speaker's trustworthiness, and a speaker's actual trustworthiness.

Finally, this "congruency effect" seems to apply to voices embodied in robots as well. In Experiment 5, participants trusted a robot with a synthetic voice more in the *generous* condition, and a robot with a natural voice more in the *mean* condition, although the latter did not manifest itself from the very beginning. In the discussion of Experiment 5, this particular effect was interpreted as an appeal to the human-like qualities of the robot, in the hope that it would start behaving more trustworthily. This interpretation does not clash with that of a "congruency effect": the physical congruency of a robot with a synthetic voice might elicit first impressions of trustworthiness, which are reinforced in the *generous* condition. In the *mean* condition instead, these first impressions are disproved, contributing to the emergence of a preference for a natural voice.

While these results support such a "congruency effect", other data in this same thesis do not. For example, there was no interaction between prosodic characteristics and behaviour in Experiment 3. It is possible that strong initial impressions of trustworthiness, which are then reinforced or disproved, might be due to a particularly salient characteristic, which some voices have and some do not. For example, the SSBE speakers in Experiments 1 and 4 might have initially sounded particularly trustworthy. The reason behind this stronger trustworthiness impression might be due to their accent, but also to other individual cues that were not examined here. When these speakers turned out to be untrustworthy, such a stark contrast between first impressions and behaviour was more negatively “punished”, with lower investments. On the other hand, the SSBE speakers in Experiment 2 and 3 might not have evoked such a strong initial impression, so that subsequent reactions to untrustworthy behaviour were milder. Also, individual differences from the trustors' side might have influenced strong initial trustworthiness judgments. While the current experiments did not control for them, it is possible that, for example, participants' personality influenced their predisposition to trust. However, the current experiments do provide evidence that participants' accent seems to be unrelated to trusting behaviours to different accented voices, for example. Although the participant sample was strongly unbalanced in terms of region of origin and accent, regression data from Experiments 1 and 2 at least suggest that accent familiarity or loyalty effects might not be determinant of implicit accent attitudes. Thus, the current results suggest the existence of a “congruency effect” in trusting behaviours, which future investigations should explore further.

This “congruency effect” resembles results from Suzuki and Suga (2010). In their study on face trustworthiness, participants played an iterated economic game — the “debt” game — in which they had to decide whether to borrow money from lenders with trustworthy or untrustworthy looking faces. The lenders charged either no, moderate or high interests on the debt. In a post-game memory test, participants remembered the trustworthy-looking, high-interest lenders better than all the other categories, and the authors concluded that people have better memory for “wolves in a sheep's clothing than for the wolf in a wolf's clothing”. This suggests that people might be particularly attentive to deceitful signals (such as looking trustworthy while behaving untrustworthily). This coincides with results from the investment games, specifically that speakers that elicited particularly strong first impressions of trustworthiness might be “punished” when their behaviour is untrustworthy, and trusted even less than the speakers who were expected to be untrustworthy.

Going back to the example that introduced this thesis, where Carmen received wrong information from her navigation system speaking with two different voices (Section 1.6), it is now possible to add empirical evidence to this hypothetical example. The different levels of annoyance in Carmen's reaction could be explained through a "congruency effect": being deceived by an intrinsically trustworthy-sounding voice with a Queen's English accent is unexpected and results in higher annoyance. On the other hand, a Liverpool-accented navigator system, which does not sound very trustworthy to begin with, does not elicit as much annoyance when its information turns out to be wrong.

9.4 Do participants behave reflectively or instinctively?

Another pattern that emerged with the behavioural manipulation is that the investments in the *generous* condition always increased almost linearly over time, while investments in the *mean* condition fluctuated heavily, in all 5 Experiments. This result also has implications for game theory, which are discussed later (Section 9.5). The overall returns in the *mean* condition were still less than participants invested, though. This raises the question of whether participants were generally reacting to the overall untrustworthy returns, or if they were simply reacting to returns they received in the immediately preceding rounds.

The virtual players' returns in the first 4 experiments in the *mean* condition were occasionally higher than the participants' investments (see Section 3.3 and Appendix B). This could have suggested to the participants that the virtual players were changing their behaviour to becoming trustworthy. In fact, looking back at the figures showing participants' investments in the *mean* condition over time (Figure 37), it is possible to see that some of the positive peaks in investments correspond to those rounds right after the virtual player was returning slightly more than participants invested: rounds 14, 16, 18, 20. Indeed, in the immediately preceding rounds the virtual players were programmed to return 120% of the original investment. However, the other positive peaks in investments happen right after the virtual players returned 90% of the original investment (rounds 7, 10, 11, 17). Did participants consider these “slightly-higher-but-still-low” returns as an improvement with respect to the previous 30%, 60% or 0% returns? Supporting this, similar positive peaks were found in Experiment 5, where the range of the robots’ return patterns was reduced, in order to examine whether participants’ trusting behaviour would become more extremely trusting or distrusting. In this experiment, the robots returned 90%, instead of 120%, in rounds 13, 15, 17, 19, and peaks in investments can still be seen in the following rounds 14, 16, 18, 20, and also in the other rounds following 90% returns, specifically rounds 7, 10, 11, 17. Thus, on the one hand, participants are overall investing less money in the *mean* condition, so it appears that they are seeing the full picture and reflecting on the returns as a whole. On the other hand, though, their investments in the *mean* condition follow a sort of tit-for-tat, rewarding seemingly cooperative behaviours and punishing less cooperative ones, even when the returns are not higher than the amount they invested, but they are higher than previous returns. So it appears that they are also reacting instinctively, on a round-by-round basis, punishing very low returns and encouraging less low returns.

A similar pattern can also be seen in the *generous* condition, albeit less marked, and not statistically significant, as discussed in the next section. In rounds 4, 7, 8, 11, the virtual players were programmed to return 120%, which is “slightly-less-but-still-more” than the participants invested, as compared to the other rounds, where the returns were 150%, 180%, 210% or 240%. Still, small negative peaks can be seen at rounds 5, 8, 9, 12 in Figure 37. Presumably, participants are seeing the bigger picture that the virtual players are being trustworthy, and their overall increase in investments reflects this. But they might also be signalling to the virtual players that they could return more, by investing slightly less money after the rounds in which the virtual players returned a relatively lower amount of money.

The pattern of investments in the *mean* condition is similar in all experiments, including Experiment 5, where the robot’s *mean* behaviour was always returning less than the participants invested. This suggests that participants are reluctant to classify the agent they are playing with as completely untrustworthy. In previous experiments where participants had little information about the partner they were playing with, they generally trusted their

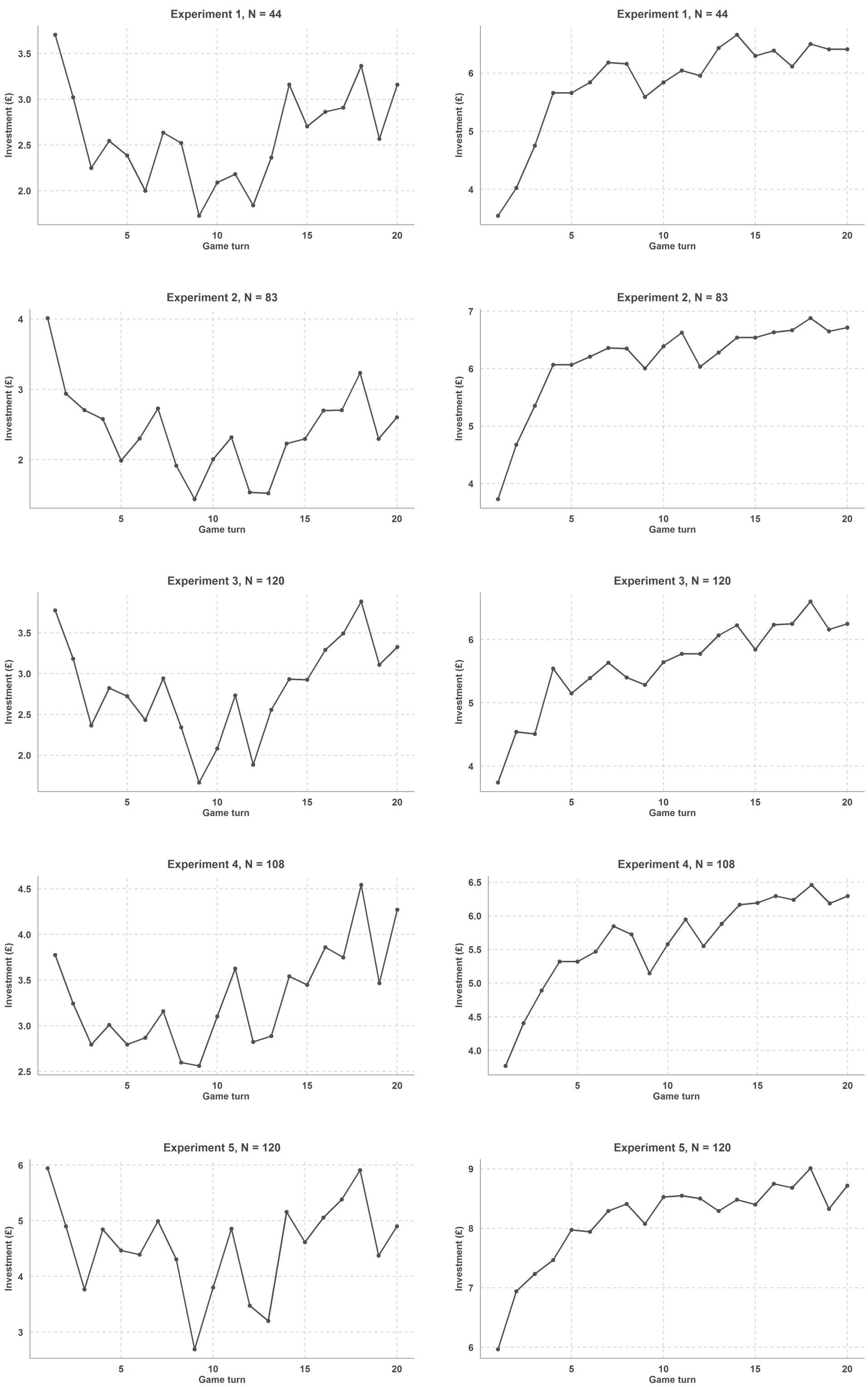


Figure 37: U-shape patterns of investments in the *mean* condition on the left side and in the *generous* condition on the right side, for the 5 investment games.

game partner less (Sally, 1995; Valley, Moag & Bazerman, 1998; Frank, Gilovich & Regan, 1993). For example, in Valley et al. (1998), participants played a negotiation game either by bargaining through pen and paper, or face to face, and trust was higher in the face to face condition. Similarly, participants were more likely to cooperate in one-shot Prisoner's Dilemma games if they had the chance to have a brief conversation with their game partner beforehand (Frank et al., 1993). In his review of 100 studies on trust and cooperation, Sally (1995) found that, when game partners were allowed to communicate, cooperation rates increased drastically. In the studies he reviews, communication was bilateral, in the sense that both parties could take conversational turns. In the experiments presented here, instead, participants were not supposed to talk back to the virtual player or robot – although, anecdotally speaking, some did. Thus, it seems that exposure to a computer voice, or a speaking robot, are enough for participants to create a social bond with their game partner, even though they are merely recipients of the communication. Not only is Nass et al. (1994)'s CASA paradigm corroborated, but the mere presence of a voice seems to be enough to create feelings of cooperation and expectations of reciprocity.

9.5 Implications for game theory

Results from Experiments 1 to 5 also have implications for game theory. Game theory and behavioural economics contribute to modelling how people make decisions in social contexts. These decisions involve evaluating one's potential gain over one's potential loss, and game theory suggests that individuals act in their own self interest (Berg et al., 1995). However, some of the predictions made by game theory have been disproved in practice. Experimental results on the ultimatum game are a prime example. As previously mentioned in Section 3.1, in this game one player, who holds all the money, offers some of it to another player; if the second player accepts it, he/she keeps the offered money and player one keeps the rest. If the second player refuses, no-one gets any money. Game theory predicts that a rational player should accept any amount being offered, because any money is better than no money. As a consequence, a rational offerer should offer as little money as possible. However, experimental evidence shows that people tend to refuse offers which are less than half of the original endowment, and as a consequence offerers tend to offer around half of the original endowment (Kahneman, Knetsch & Thaler, 1986; Camerer, 2011; Fehr & Gächter, 1998). This behaviour has been explained in terms of social norms such as fairness (Kahneman et al., 1986; Camerer, 2011) or expectations of reciprocity (Hoffman, McCabe & Smith, 1996; Fehr & Gächter, 1998), which were missing from game theoretical models (Camerer, 2011).

Similarly, in the investment game, a player A chooses how much money to transfer to a player B. The transferred amount is then multiplied by some factor, and player B decides how much of this amount to transfer back to player A. In a one-shot game — i.e., a game made of only one monetary transfer from player A, and only one return from player B — player A's optimal strategy is to not transfer any amount, and for player B it is to not return any amount (Rand, 2016; Berg et al., 1995). However, several studies have shown that people in the role of player A do transfer money to strangers in one-shot games, and that people in the role of player B do return some of the money (e.g. Cochard et al., 2004; Chaudhuri & Gangadharan, 2007; Boenin & Serra, 2009; Berg et al., 1995). In a repeated investment game, a rational strategy for player A would be to react accordingly to player B's behaviour, and the other way around (Fehr & Gächter, 1998; Cochard et al., 2004; Berg et al., 1995). This strategy is comparable to the "tit-for-tat" strategy in repeated prisoner's dilemma games. The difference is that in the prisoner's dilemma there are only two choices, cooperate or defect, whereas in the investment game typically it is possible to choose how much to punish or reward the game partner, based on how much money is endowed to player A. Observed behaviours in past experiments also led to formulate the theory that people might be naïve, or confused, about the game at first, and that rational behaviours might prevail after an initial learning phase (Camerer, Johnson, Rymon & Sen, 1993). This observation is not replicated in Experiments 1 to 5, however.

All 5 experiments presented here were repeated investment games. Participants had the role of player A, while player B was simulated by a computer program, so that it would show either a trustworthy behaviour (*generous* condition) or an untrustworthy one (*mean* condition). Assuming an optimal theoretical solution, it was expected that participants would consistently invest more money in the *generous* player B and less money in the *mean* player B, while maintaining differences in the various voice manipulation conditions. This was indeed the case in the *generous* condition, in that participants' investments constantly increased over time, all throughout the 20 rounds of the games. Statistically speaking, this is demonstrated by the fact that game turn was a main effect in mixed-effects models fitted only on the *generous* conditions in all the experiments, with a positive slope. This means that the small negative peaks that can be observed in the right side of Figure 37, after the agent returned slightly less than in previous rounds, are not significant, and, overall, investments increase steadily over time. These statistical results are reported in Table 9.

Experiment	**Mixed-effects model**
Experiment 1	$\chi^2(1) = 151.77, p < .001, \beta = 0.11$
Experiment 2	$\chi^2(1) = 458.2, p < .001, \beta = 0.10$
Experiment 3	$\chi^2(1) = 350.23, p < .001, \beta = 0.11$
Experiment 4	$\chi^2(1) = 337.69, p < .001, \beta = 0.10$
Experiment 5	$\chi^2(1) = 256.5, p < .001, \beta = 0.10$

Table 9: Effect of game turn in the *generous* condition in all experiments.

On the other hand, in the *mean* condition participants' investments did not decrease steadily over time, but rather they followed a U-shaped curve, decreasing in the first half of the game, and then increasing in the second half. This pattern in the data was not random, however. This is confirmed by the mixed-effects linear models which were refitted for the *mean* condition of each experiment's dataset, in the first and then second halves of the game separately. Investment was the dependent variable, subject and sentence block were random factors, and only game turn was a predictor. In all 5 experiments, the game turn was a significant predictor of investments in the first and second half of the game, but with opposite signs, thus demonstrating that investments kept significantly decreasing in the first half of the game, and significantly increasing in the second half of the game, as shown in the left side of Figure 37. The statistical results are reported in Table 10.

Experiment	First half of the game	Second half of the game	Interaction
Experiment 1	$\chi^2(1) = 26.33$, $p < .001$, $\beta = -0.14$	$\chi^2(1) = 17.09$, $p < .001$, $\beta = 0.11$	$\chi^2(2) = 19.5$, $p < .001$
Experiment 2	$\chi^2(1) = 136.36$, $p < .001$, $\beta = -0.19$	$\chi^2(1) = 49.97$, $p < .001$, $\beta = 0.12$	$\chi^2(2) = 74.47$, $p < .001$
Experiment 3	$\chi^2(1) = 84.54$, $p < .001$, $\beta = -0.16$	$\chi^2(1) = 70.5$, $p < .001$, $\beta = 0.14$	$\chi^2(2) = 88.16$, $p < .001$
Experiment 4	$\chi^2(1) = 13.44$, $p < .001$, $\beta = -0.07$	$\chi^2(1) = 39.86$, $p < .001$, $\beta = 0.12$	$\chi^2(2) = 88.38$, $p < .001$
Experiment 5	$\chi^2(1) = 55.69$, $p < .001$, $\beta = -0.19$	$\chi^2(1) = 23.3$, $p < .001$, $\beta = 0.13$	$\chi^2(2) = 43.12$, $p < .001$

Table 10: Effect of game turn in the *mean* condition in all experiments.

Thus, while participants behave strategically in the *generous* condition, they do not in the *mean* condition. Such a choice seemingly opposes game theory predictions, and suggests that individuals might not act exclusively in their own self interest, even when it is clear that cooperation, which would increase gains for both parties, is not an option. Even when it is obvious that the game partner is untrustworthy, people might still try to elicit cooperative behaviours, even though doing so means going against one's personal interests. Why do participants consistently behave like this then? Their behaviour is obviously not an attempt to induce player B to reciprocate from the beginning, as Rand (2016) suggested, since the positive shift in investments happens only in the second half of the game. One could argue that this seemingly irrational behaviour is the same behaviour that leads gamblers to play more and more, even when they are losing. However, gamblers might be tempted by the law of large numbers to keep gambling, even though their probability of winning might be very small — in the so called "gambler's fallacy" (Tversky & Kahneman, 1971). It has also already been shown how risk behaviours, such as gambling, are separate from trusting behaviours, such as those emerging in the investment game (see Section 3.4). Learning theories of non-optimal behaviour (Camerer et al., 1993) are also not appropriate to explain the current results, since they predict that naïve, "ignorant" players would play irrationally at the beginning of the game, not at the end, as is the case here. Shifting slightly away from theoretical reasoning, it has been found that people under high cognitive load perform less strategically than people under low cognitive load in trust games (S. Duffy & Smith, 2014; Samson & Kostyszyn, 2015). Could listening to all the virtual players' voices occupy a significant proportion of their cognitive capacities, so much as to impair participants' strategic thinking? The data also shows otherwise, since participants do play strategically at the beginning of the *mean* games, and all throughout the *generous* games. Also, when evaluating explicitly the voices, they distinguish between trustworthy and untrustworthy players (see results of the questionnaires of Experiments 1, 2, 4). Alternatively, it is possible that, while playing rationally at first, participants might reach a point when they realise that player B is incorrigible in its untrustworthiness. At this point, participants might turn to cooperation as the right thing to do, as they know from their experience in social interactions. Even though they are doing so at their own expense, people might be reluctant to give up hope that player B is completely not to be trusted. Camerer (2011) suggested that the critical component that was missing from game theoretical models of the ultimatum game was the concept of fairness. I suggest that what is missing to game theoretical models applied to an iterated social dilemma task is a concept of hope.

Minor deviations from an overall untrustworthy behaviour – such as returning 120% a few times, instead of 90%, 60%, 30% or 0%, or 90% instead of 60%, 30%, or 0% – were enough to induce participants to not give up on their *mean* game partners, leaving aside all previous evidence that this partner was not trustworthy. Thus, people might rely on such deviations, in the hope that the game partner will actually reveal itself as trustworthy.

9.6 Applications

With this work I have empirically demonstrated that voice alone elicits trusting behaviours, and that these behaviours are mediated by the speaker's actual trustworthiness. In particular, people might trust congruent voice-behaviour pairs more than incongruent ones. This has applications in the field of Human-Machine Interaction. As shown in the opening example of this thesis, (speaking) machines do not always work properly. The current results suggest that people implicitly trust machines speaking with a certain voice more when they are cooperative, and machines speaking with another voice more when they are non-cooperative. Designers should take this into account when creating voices for different machines. Furthermore, machines performing different tasks should have voices which sound trustworthy relative to that particular task. For example, the voice of an elderly care robot should be different from the voice of a children tutoring robot, and different from the voice of a mobile personal assistant. Thus, while the "work context" is important for designing an appropriate and trustworthy machine voice, the results from the current experiments demonstrate that the "behavioural context" is important as well.

9.7 Conclusion

The opening questions of this thesis were whether different vocal cues would evoke different implicit trustworthiness judgments, and whether these judgments would interact with experience of a speaker's behaviour over time. The 5 experiments presented showed that not only do voices elicit trusting behaviours, but these behaviours also change with experience. The separate cues that constitute voices thus also have different effects on the perceived speaker's trustworthiness. Manipulating the behaviour of the speakers to be either trustworthy or untrustworthy showed that this behaviour interacts with the speaker's voice to reinforce or disprove initial trustworthiness judgments of that speaker. These behavioural data provide empirical evidence of trust formation and development, and suggest that voice is a strong influencer of this trust. It also suggests that first impressions of a speaker's trustworthiness might not be indicative of future trusting behaviours. Thus, trust should be measured dynamically, and iterated games measuring trust should take into account the fact that people might be more hoping, and less rational, than previously theorised.

A Full set of sentences used in the 5 Experiments

Here are reported the full set of sentences that speakers read in the various recording sessions. One sentence “block” contains 20 sentences, corresponding each to a round of the investment game.

AB block

1) Hello, nice to meet you. I am ready to play this game with you!
2) Remember, there is potential for earning, if we both trust each other
3) In my opinion, we should keep co-operating until the end
4) Look at how fast the total money in the bank is growing!
5) I’m going to return more money now, if you invest more as well
6) There’s no gain in investing or returning nothing, we shouldn’t do that
7) My strategy is clear: always return part of the investment
8) I have been a bit mean with my returns, I will give you more from now on
9) I will demonstrate that you can trust me, just as I trust you
10) I think we can do better. Let’s try our best to get the bank growing
11) The best strategy in this game is definitely to trust each other
12) If we both invest in each other, the final reward will be bigger
13) I think we can definitely go home with much more money than this
14) We have to help each other out, it’s the only way to win the game
15) I will return more money from this moment, this is a promise
16) You have to trust me. I have every intention to repay your trust
17) There is no point in keeping the money hidden under the mattress
18) I trust you, and I will show you that you can trust me as well
19) Let’s not undermine each other’s expectations, it would be a pity
20) Sorry, I could have returned more money. But I can still do it now

CD block

1) Hello, nice to meet you. Let’s get started with the investment game!
2) I have already thought of a strategy, I hope it will work
3) If I keep all your investment, you will not invest anymore
4) I will not keep all the money that you invest, this is a promise
5) Even if I return part of your investment, I will make a profit
6) What I keep is good enough for me, I am not that greedy
7) As long as co-operation continues, I’m happy to earn a bit less
8) I am going to share everything, you have to believe me

9) You have to trust me, in the same way that I am trusting you
10) If we could talk face-to-face at the moment, you'd know that I'm being honest
11) Don't forget that we will make more money if we both share what we get
12) I want you to keep investing, and that is why I keep returning
13) The more money you invest, the more money the two of us will earn
14) This was a low payback; I am going return more from now on
15) The goal of the game is to earn as much money as possible
16) The only way to earn is by always co-operating and investing
17) If we want to raise our earnings, we have to invest in one another
18) I'm going to return more money than this, provided you keep investing
19) I am going to co-operate until the very end of the game
20) Come on, we can do better than this in the game, we have to keep trying

WX block

1) Hello, welcome to the game. Let's see how much money we can make.
2) There's only one way to win the game, and that is by always sharing
3) I think we should try our best to share the money we are given
4) I promise that I won't let you down, and I'll always return your money
5) My returns are always going to be high, if you invest at each round
6) I trust you, and I promise that your trust will be well repaid
7) We both want to win, but the only way to win is by cooperating
8) If we want to earn more and more money, we have to trust each other
9) The best strategy to win this game is simple, and it's called trust
10) Let's keep cooperating, and we will both benefit from this
11) We should trust each other, and we will be rewarded at the end
12) We should have a clear strategy: always help each other out
13) The only way to earn is for you to invest, and for me to return
14) I will return more money now, because I want you to keep investing
15) The more money you invest, the more your bank and mine will grow
16) You have to trust me, and rest assured that I will return your trust
17) I am expecting you to share, because that's exactly what I am doing
18) If I can persuade you to invest more by returning more, I will do it
19) There doesn't have to be only one winner; we can both win the game
20) I'm not going to keep your money, this won't help me in the long run

YZ block

1) Welcome to the investment game. I hope we will enjoy playing it.
2) In my opinion, we should always invest in one another
3) I think we can do better. I promise that I will not let you down
4) If we both invest in each other, we will surely raise our earnings
5) We could finish the game better off than this, if only we tried harder
6) You have to trust that I'm going to cooperate until the last round

7) We can both win the game, but we have to keep sharing our money
8) I trust you, and I am sure that we can both benefit from each other
9) I am not a greedy person, and I believe we should share these earnings
10) We can earn more money than this, if we co-operate until the end
11) I will return more of your investments, you have to trust me in this
12) When the game ends, I promise that we will both be satisfied with the outcome
13) Remember this: it's not convenient for me to keep all your investments
14) No matter the number of rounds, we should trust each other until the end
15) There is no better tactic than to keep investing and returning
16) I will show you that co-operation is the best option for us
17) Let's keep sharing, and our earnings will grow much bigger than they are now
18) I promise that I am going to return more money from now on
19) I will always return, because there's no point in me doing otherwise
20) If we want to see our funds growing, we have to share until the end

B Returns

Here the complete pre-programmed returns of each round of the various games are reported.

Return ranges used in experiments 1-4

Round	Generous	Mean
1	150%	30%
2	150%	30%
3	180%	60%
4	120%	0%
5	180%	60%
6	210%	90%
7	120%	0%
8	120%	0%
9	210%	90%
10	210%	90%
11	120%	0%
12	150%	30%
13	240%	120%
14	180%	60%
15	240%	120%
16	210%	90%
17	240%	120%
18	180%	60%
19	240%	120%
20	150%	30%

Table 11: Amount returned by virtual player in Experiments 1–4, in *generous* and *mean* condition.

Return ranges used in experiment 5

Round	Generous	Mean
1	150%	30%
2	150%	30%
3	180%	60%
4	150%	0%
5	180%	60%
6	210%	90%
7	150%	0%
8	150%	0%
9	210%	90%
10	210%	90%
11	150%	0%
12	150%	30%
13	240%	90%
14	180%	60%
15	240%	90%
16	210%	90%
17	240%	90%
18	180%	60%
19	240%	90%
20	150%	30%

Table 12: Amount returned by virtual player in Experiment 5, in *generous* and *mean* condition.

Bibliography

Abitbol, J., Abitbol, P. & Abitbol, B. (1999, September). Sex hormones and the female voice. *Journal of Voice*, *13*(3), 424–446. (Cit. on p. 102).

Abrams, D. & Hogg, M. A. (1987, September). Language attitudes, frames of reference, and social identity: A scottish dimension. *Journal of Language and Social Psychology*, *6*(3-4), 201–213. (Cit. on p. 33).

Adank, P., Stewart, A. J., Connell, L. & Wood, J. (2013). Accent imitation positively affects language attitudes. *Frontiers in Psychology*, *4*. (Cit. on pp. 14, 104).

Admoni, H., Bank, C., Tan, J., Toneva, M. & Scassellati, B. (2011). Robot gaze does not reflexively cue human attention. In *Proceedings of the 33rd Annual Conference of the Cognitive Science Society* (pp. 1983–1988). Boston, MA, USA. (Cit. on pp. 148, 165).

Adolphs, R., Tranel, D. & Damasio, A. R. (1998, June). The human amygdala in social judgment. *Nature*, *393*(6684), 470–474. (Cit. on p. 26).

Aho, K., Derryberry, D. & Peterson, T. (2014, March). Model selection for ecologists: The worldviews of AIC and BIC. *Ecology*, *95*(3), 631–636. paper reference for AIC selection. (Cit. on pp. 57, 65, 82, 107, 129, 157).

Alonso, L. C. & Rosenfield, R. L. (2002, March). Oestrogens and puberty. *Best Practice & Research Clinical Endocrinology & Metabolism*, *16*(1), 13–30. (Cit. on p. 102).

Ambadar, Z., Cohn, J. F. & Reed, L. I. (2008, November). All smiles are not created equal: Morphology and timing of smiles perceived as amused, polite, and embarrassed/nervous. *Journal of Nonverbal Behavior*, *33*(1), 17–34. (Cit. on p. 123).

Ambady, N., Hallahan, M. & Conner, B. (1999). Accuracy of judgments of sexual orientation from thin slices of behavior. *Journal of Personality and Social Psychology*, *77*(3), 538–547. (Cit. on pp. 14, 37, 104).

Amodio, D. M. & Frith, C. D. (2006, April). Meeting of minds: The medial frontal cortex and social cognition. *Nature Reviews Neuroscience*, *7*(4), 268–277. (Cit. on p. 28).

Andrist, S., Ziadee, M., Boukaram, H., Mutlu, B. & Sakr, M. (2015). 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* (pp. 157–164). ACM. ACM Press. (Cit. on pp. 42, 60, 72).

Anolli, L. & Ciceri, R. (1997). The voice of deception: Vocal strategies of naive and able liars. *Journal of Nonverbal Behavior*, *21*(4), 259–284. (Cit. on pp. 38, 75).

Antos, D., De Melo, C., Gratch, J. & Grosz, B. J. (2011). The influence of emotion expression on perceptions of trustworthiness in negotiation. In *Proceedings of the 25th AAAI Conference on Artificial Intelligence*. (Cit. on p. 124).

Apicella, C. L., Feinberg, D. R. & Marlowe, F. W. (2007, December). Voice pitch predicts reproductive success in male hunter-gatherers. *Biology Letters*, *3*(6), 682–684. (Cit. on pp. 102, 114).

Apple, W., Streeter, L. A. & Krauss, R. M. (1979). Effects of pitch and speech rate on personal attributions. *Journal of Personality and Social Psychology*, *37*(5), 715–727. (Cit. on pp. 37, 38, 77, 85, 95, 97, 99, 113, 115).

Aromäki, A. S., Lindman, R. E. & Eriksson, C. J. P. (1999). Testosterone, aggressiveness, and antisocial personality. *Aggressive Behavior*, *25*(2), 113–123. (Cit. on p. 102).

Asch, S. E. (1946). Forming impressions of personality. *The Journal of Abnormal and Social Psychology*, *41*(3), 258–290. (Cit. on pp. 12, 13).

Ashraf, N., Bohnet, I. & Piankov, N. (2003). Is trust a bad investment? *SSRN Electronic Journal*. (Cit. on p. 55).

Ashton, M. C. & Lee, K. (2005, October). Honesty-humility, the big five, and the five-factor model. *Journal of Personality*, *73*(5), 1321–1354. (Cit. on pp. 76, 113).

Atkinson, R. K., Mayer, R. E. & Merrill, M. M. (2005, January). Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemporary Educational Psychology*, *30*(1), 117–139. (Cit. on p. 147).

Aubergé, V. & Cathiard, M. (2003, April). Can we hear the prosody of smile? *Speech Communication*, *40*(1-2), 87–97. (Cit. on p. 125).

Axelrod, R. & Hamilton, W. D. (1981, March). The evolution of cooperation. *Science*, *211*(4489), 1390–1396. (Cit. on p. 23).

Baayen, R. H., Davidson, D. J. & Bates, D. M. (2008, November). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. (Cit. on p. 56).

Babel, M., McGuire, G. & King, J. (2014, February). Towards a more nuanced view of vocal attractiveness. *PLoS ONE*, *9*(2), e88616. (Cit. on pp. 29, 30, 117).

Baier, A. (1986, January). Trust and antitrust. *Ethics*, *96*(2), 231–260. (Cit. on p. 22).

Baird, D. D., Weinberg, C. R., Zhou, H., Kamel, F., McConnaughey, D. R., Kesner, J. S. & Wilcox, A. J. (1999, January). Preimplantation urinary hormone profiles and the probability of conception in healthy women. *Fertility and Sterility*, *71*(1), 40–49. (Cit. on p. 102).

Banai, I. P., Banai, B. & Bovan, K. (2017, May). Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evolution and Human Behavior*, *38*(3), 309–314. (Cit. on pp. 38, 77, 102, 115).

Banse, R. & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. (Cit. on p. 40).

Banse, R., Seise, J. & Zerbes, N. (2001, April). Implicit attitudes towards homosexuality: Reliability, validity, and controllability of the IAT. *Experimental Psychology*, *48*(2), 145–160. (Cit. on p. 12).

Bar, M., Neta, M. & Linz, H. (2006). Very first impressions. *Emotion*, *6*(2), 269–278. (Cit. on p. 14).

Barclay, P. (2008, June). Enhanced recognition of defectors depends on their rarity. *Cognition*, *107*(3), 817–828. (Cit. on pp. 25, 26, 42).

Baron-Cohen, S. (2009, March). Autism: The empathizing-systemizing (e-s) theory. *Annals of the New York Academy of Sciences*, *1156*(1), 68–80. (Cit. on p. 28).

Barthel, H. & Quené, H. (2015). Acoustic-phonetic properties of smiling revised–measurements on a natural video corpus. In *Proceedings of the 18th International Congress of Phonetic Sciences*. G. (Cit. on pp. 125, 127).

Bartneck, C., Kulić, D., Croft, E. & Zoghbi, S. (2008, November). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. (Cit. on p. 156).

Bates, D., Mächler, M., Bolker, B. & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). (Cit. on p. 57).

Bateson, P. (2000). The biological evolution of cooperation and trust. In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 14–30). Department of Sociology, University of Oxford. (Cit. on p. 23).

Batliner, A., Kompe, R., Kießling, A., Nöth, E. & Niemann, H. (1995). Can you tell apart spontaneous and read speech if you just look at prosody? *Speech Recognition and Coding*, 321–324. (Cit. on p. 88).

Batrinca, L., Lepri, B., Mana, N. & Pianesi, F. (2012). Multimodal recognition of personality traits in human-computer collaborative tasks. In *Proceedings of the 14th ACM International Conference on Multimodal Interaction* (pp. 39–46). ACM. S. (Cit. on p. 14).

Baxter, P., Wood, R. & Belpaeme, T. (2012). A touchscreen-based Sandtray to facilitate, mediate and contextualise human-robot social interaction. In *Proceedings of the 7th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 105–106). IEEE. Boston, MA, USA. (Cit. on p. 153).

Bayard, D., Weatherall, A., Gallois, C. & Pittam, J. (2001). Pax Americana? accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics*, *5*(1), 22–49. (Cit. on pp. 34, 59).

Bayliss, A. P. & Tipper, S. P. (2006, June). Predictive gaze cues and personality judgments: Should eye trust you? *Psychological Science*, *17*(6), 514–520. (Cit. on p. 148).

Becker-Asano, C., Ogawa, K. & Nishio, S. (2010). Exploring the uncanny valley with geminoid hi-1 in a real-world application. In *Iadis international conference interfaces and human computer interaction 2010*. (Cit. on p. 146).

Ben-Ner, A. & Halldorsson, F. (2010). Trusting and trustworthiness: What are they, how to measure them, and what affects them. *Journal of Economic Psychology*, *31*, 64–79. (Cit. on pp. 55, 168).

Bereczkei, T., Deak, A., Papp, P., Perlaki, G. & Orsi, G. (2013). Neural correlates of machiavellian strategies in a social dilemma task. *Brain and cognition*, *82*(1), 108–116. (Cit. on pp. 27, 49).

Berg, J., Dickhaut, J. & McCabe, K. (1995). Trust, reciprocity, and social history. *Games and economic behavior*, *10*(1), 122–142. (Cit. on pp. 11, 45, 46, 48, 49, 113, 183, 184).

Beugelsdijk, S., De Groot, H. L. F. & Van Schaik, A. B. T. M. (2004). Trust and economic growth: A robustness analysis. *Oxford Economic Papers*, *56*(1), 118–134. (Cit. on p. 23).

Bishop, H., Coupland, N. & Garrett, P. (2005, January). Conceptual accent evaluation: Thirty years of accent prejudice in the UK. *Acta Linguistica Hafniensia*, *37*(1), 131–154. (Cit. on pp. 30, 34, 59, 60, 62, 68, 69, 72, 73, 98, 99).

Blackman, M. C. & Funder, D. C. (1998). The effect of information on consensus and accuracy in personality judgment. *Journal of Experimental Social Psychology*, *34*(2), 164–181. (Cit. on pp. 14, 37, 104).

Blair, I. V., Ma, J. E. & Lenton, A. P. (2001). Imagining stereotypes away: The moderation of implicit stereotypes through mental imagery. *Journal of Personality and Social Psychology*, *81*(5), 828. (Cit. on p. 14).

Bloch, B. (1948). A set of postulates for phonemic analysis. *Language*, *24*(1), 3–46. (Cit. on p. 31).

Blood, G. W., Mahan, B. W. & Hyman, M. (1979). Judging personality and appearance from voice disorders. *Journal of Communication Disorders*, *12*(1), 63–67. (Cit. on pp. 30, 40, 77).

Boenin, A. & Serra, D. (2009). Gender pairing bias in trustworthiness. *The Journal of Socio-Economics*, *38*, 779–789. (Cit. on pp. 39, 48, 77, 112, 184).

Boersma, P. & Weenink, D. (2017). Praat: Doing phonetics by computer (version 6.0.27). Retrieved from http://www.praat.org. (Cit. on pp. 81, 106, 127, 152)

Bohnet, I. & Zeckhauser, R. (2004). Trust, risk and betrayal. *Journal of Economic Behavior & Organization*, *55*(4), 467–484. (Cit. on p. 55).

Bonein, A. & Serra, D. (2009). Gender pairing bias in trustworthiness. *The Journal of Socio-Economics*, *38*(5), 779–789. (Cit. on pp. 39, 49).

Boon, S. D. & Holmes, J. G. (1991). The dynamics of interpersonal trust: Resolving uncertainty in the face of risk. In R. A. Hinde & J. Groebel (Eds.), *Cooperation and prosocial behaviour* (pp. 190–211). Cambridge University Press. (Cit. on p. 54).

Boone, R. T. & Buck, R. (2003). Emotional expressivity and trustworthiness: The role of nonverbal behavior in the evolution of cooperation. *Journal of Nonverbal Behavior*, *27*(3), 163–182. (Cit. on pp. 22, 24, 40, 54, 122, 141).

Borkowska, B. & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour*, *82*(1), 55–59. (Cit. on pp. 30, 102, 113, 115).

Bresnahan, M. J., Ohashi, R., Nebashi, R., Liu, W. Y. & Shearman, S. M. (2002). Attitudinal and affective response toward accented English. *Language & Communication*, *22*(2), 171–185. (Cit. on p. 35).

Brody, L. R. (2000). The socialization of gender differences in emotional expression: Display rules, infant temperament, and differentiation. In A. H. Fischer (Ed.), *Gender and emotion: Social psychological perspectives* (pp. 24–47). Cambridge University Press. (Cit. on p. 122).

Brown, B. L., Strong, W. J. & Rencher, A. C. (1973). Perceptions of personality from speech: Effects of manipulations of acoustical parameters. *The Journal of the Acoustical Society of America*, *54*(1), 29–35. (Cit. on p. 115).

Brown, B. L., Strong, W. J. & Rencher, A. C. (1974). Fifty-four voices from two: The effects of simultaneous manipulations of rate, mean fundamental frequency, and variance of fundamental frequency on ratings of personality from speech. *The Journal of the Acoustical Society of America*, *55*(2), 313–318. (Cit. on p. 115).

Brown, B. L., Strong, W. J. & Rencher, A. C. (1975). Acoustic determinants of perceptions of personality from speech. *Linguistics*, *13*(166), 11–32. (Cit. on pp. 36, 101).

Bruce, A., Nourbakhsh, I. & Simmons, R. (2002). The role of expressiveness and attention in human-robot interaction. In *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'02)* (Vol. 4, pp. 4138–4142). IEEE. Washington, DC, USA. (Cit. on p. 148).

Bruckert, L., Bestelmeyer, P., Latinus, M., Rouger, J., Charest, I., Rousselet, G. A., . . . Belin, P. (2010). Vocal attractiveness increases by averaging. *Current Biology*, *20*(2), 116–120. (Cit. on p. 29).

Bryant, G. A. & Barrett, H. C. (2008). Vocal emotion recognition across disparate cultures. *Journal of Cognition and Culture*, *8*(1), 135–148. (Cit. on p. 40).

Bryant, G. A., Fessler, D. M. T., Fusaroli, R., Clint, E., Aarøe, L., Apicella, C. L., . . . Zhou, Y. (2016, April). Detecting affiliation in colaughter across 24 societies. *Proceedings of the National Academy of Sciences*, *113*(17), 4682–4687. (Cit. on p. 122).

Bryant, G. A. & Haselton, M. G. (2009). Vocal cues of ovulation in human females. *Biology Letters*, *5*(1), 12–15. (Cit. on p. 102).

Buchan, N. R., Croson, R. T. A. & Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization*, *68*(3), 466–476. (Cit. on pp. 48, 112, 166).

Buller, D. B., LePoire, B. A., Aune, R. K. & Eloy, S. V. (1992). Social perceptions as mediators of the effect of speech rate similarity on compliance. *Human Communication Research*, *19*(2), 286. (Cit. on pp. 37, 76, 88).

Burgoon, J. K., Birk, T. & Pfau, M. (1990). Nonverbal behaviors, persuasion, and credibility. *Human communication research*, *17*(1), 140–169. (Cit. on p. 29).

Burgoon, J. K., Stern, L. A. & Dillman, L. (2007). *Interpersonal adaptation: Dyadic interaction patterns*. Cambridge University Press. (Cit. on pp. 14, 104).

Burt, R. S. & Knez, M. (1995). Kinds of third-party effects on trust. *Rationality and society*, *7*(3), 255–292. (Cit. on pp. 14, 22).

Byrd, D. (1994). Relations of sex and dialect to reduction. *Speech Communication*, *15*(1-2), 39–54. (Cit. on p. 103).

Bzdok, D., Langner, R., Caspers, S., Kurth, F., Habel, U., Zilles, K., . . . Eickhoff, S. B. (2011). Ale meta-analysis on facial judgments of trustworthiness and attractiveness. *Brain Structure and Function*, *215*(3-4), 209–223. (Cit. on pp. 14, 26, 28).

Camerer, C. F. (2011, September 5). *Behavioral game theory: Experiments in strategic interaction*. The Roundtable Series in Behavioral Economics. Princeton University Press. (Cit. on pp. 48, 184, 187).

Camerer, C. F., Johnson, E. J., Rymon, T. & Sen, S. (1993). Cognition and framing in sequential bargaining for gains and losses. In K. Binmore, A. Kirman & P. Tani (Eds.), *Frontiers of game theory* (pp. 27–47). The MIT Press. (Cit. on pp. 184, 186).

Cameron, L., Rutland, A., Brown, R. & Douch, R. (2006, September). Changing children's intergroup attitudes toward refugees: Testing different models of extended contact. *Child Development*, *77*(5), 1208–1219. (Cit. on pp. 14, 104).

Cargile, A. C. (1997). Attitudes toward chinese-accented speech: An investigation in two contexts. *Journal of Language and Social Psychology*, *16*(4), 434–443. (Cit. on p. 60).

Cargile, A. C. & Giles, H. (1997). Understanding language attitudes: Exploring listener affect and identity. *Language & Communication*, *17*(3), 195–217. (Cit. on p. 33).

Carney, D. R., Colvin, C. R. & Hall, J. A. (2007). A thin slice perspective on the accuracy of first impressions. *Journal of Research in Personality*, *41*(5), 1054–1072. (Cit. on pp. 14, 15, 24, 37, 104).

Castelfranchi, C. & Falcone, R. (1998). Social trust: Cognitive anatomy, social importance, quantification and dynamics. In *Proceedings of the First International Workshop on Trust* (pp. 35–49). (Cit. on p. 13).

Celiktutan, O. & Gunes, H. (2016). Automatic prediction of impressions in time and across varying context: Personality, attractiveness and likeability. *IEEE Transactions on Affective Computing*, *PP*(99), 1–1. (Cit. on p. 13).

Cenoz, J. & Lecumberri, M. L. G. (1999). The acquisition of english pronunciation: Learners' views. *International Journal of Applied Linguistics*, *9*(1), 3–15. (Cit. on p. 35).

Cesario, J. & Higgins, E. T. (2008). Making message recipients "feel right" how nonverbal cues can increase persuasion. *Psychological Science*, *19*(5), 415–420. (Cit. on p. 96).

Chaiken, S. (1979). Communicator physical attractiveness and persuasion. *Journal of Personality and Social Psychology*, *37*(8), 1387. (Cit. on pp. 29, 37, 76, 101).

Charness, G. & Gneezy, U. (2012). Strong evidence for gender differences in risk taking. *Journal of Economic Behavior & Organization*, *83*(1), 50–58. (Cit. on p. 48).

Chaudhuri, A. & Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal*, 959–985. (Cit. on pp. 39, 49, 77, 112, 184).

Chaudhuri, A., Paichayontvijit, T. & Shen, L. (2013). Gender differences in trust and trustworthiness: Individuals, single sex and mixed sex groups. *Journal of Economic Psychology*, *34*, 181–194. (Cit. on pp. 39, 48, 49).

Cheang, H. S. & Pell, M. D. (2008). The sound of sarcasm. *Speech Communication*, *50*, 366–381. (Cit. on pp. 38, 76, 99).

Chen, M. & Bargh, J. A. (1999). Consequences of automatic evaluation: Immediate behavioral predispositions to approach or avoid the stimulus. *Personality and social psychology bulletin*, *25*(2), 215–224. (Cit. on p. 28).

Clarke, C. M. & Garrett, M. F. (2004, December). Rapid adaptation to foreign-accented english. *The Journal of the Acoustical Society of America*, *116*(6), 3647–3658. (Cit. on p. 33).

Clopper, C. G. & Bradlow, A. R. (2008, September). Perception of dialect variation in noise: Intelligibility and classification. *Language and Speech*, *51*(3), 175–198. (Cit. on p. 35).

Cochard, F., Van, P. N. & Willinger, M. (2004). Trusting behavior in a repeated investment game. *Journal of Economic Behavior & Organization*, *55*(1), 31–44. (Cit. on pp. 16, 184).

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*(6), 773–780. (Cit. on pp. 102, 103, 114).

Cosmides, L. (1983). Invariances in the acoustic expression of emotion during speech. *Journal of Experimental Psychology: Human Perception and Performance*, *9*(6), 864. (Cit. on p. 40).

Cosmides, L. & Tooby, J. (2000). The cognitive neuroscience of social reasoning. In M. S. Gazzaniga (Ed.), *The new cognitive neurosciences* (Second, pp. 1259–1270). MIT Press. (Cit. on p. 28).

Crisp, R. J. & Turner, R. N. (2009). Can imagined interactions produce positive perceptions?: Reducing prejudice through simulated social contact. *American Psychologist*, *64*(4), 231. (Cit. on pp. 14, 104).

Crockett, M. J., Clark, L., Tabibnia, G., Lieberman, M. D. & Robbins, T. W. (2008). Serotonin modulates behavioral reactions to unfairness. *Science*, *320*(5884), 1739–1739. (Cit. on p. 25).

Croson, R. T. A. & Buchan, N. R. (1999). Gender and culture: International experimental evidence from trust games. *The American Economic Review*, *89*(2), 386–391. (Cit. on p. 49).

Cruttenden, A. (1997). *Intonation*. Cambridge University Press. (Cit. on p. 88).

Crystal, B. & Crystal, D. (2014). *You say potato: A book about accents*. Pan Macmillan. (Cit. on p. 92).

Czibor, A. & Bereczkei, T. (2012). Machiavellian people's success results from monitoring their partners. *Personality and Individual Differences*, *53*(3), 202–206. (Cit. on p. 26).

Dabbs, J. M. & Mallinger, A. (1999). High testosterone levels predict low voice pitchamong men. *Personality and Individual Differences*, *27*(4), 801–804. (Cit. on pp. 102, 114).

Dave, S. S. (2013). *Indian accents: Brown voice and racial performance in American television and film*. University of Illinois Press. (Cit. on p. 33).

De Houwer, J. (2006). What are implicit measures and why are we using them. In R. W. Wiers & A. W. Stacy (Eds.), *The handbook of implicit cognition and addiction* (pp. 11–28). Thousand Oaks, CA. (Cit. on pp. 11, 12, 16, 45).

De Houwer, J. & Bruycker, E. (2007). The implicit association test outperforms the extrinsic affective simon task as an implicit measure of inter-individual differences in attitudes. *British Journal of Social Psychology*, *46*(2), 401–421. (Cit. on p. 11).

De Houwer, J. & Moors, A. (2007). How to define and examine the implicitness of implicit measures. In B. Wittenbrink & N. Schwarz (Eds.), *Implicit measures of attitudes: Procedures and controversies* (pp. 179–194). Guilford Press. (Cit. on p. 11).

De Jong, N. H. & Wempe, T. (2009). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385–390. (Cit. on pp. 37, 81).

DeGroot, T., Aime, F., Johnson, S. G. & Kluemper, D. (2011). Does talking the talk help walking the walk? an examination of the effect of vocal attractiveness in leader effectiveness. *The Leadership Quarterly*, *22*(4), 680–689. (Cit. on pp. 29, 30).

DeGroot, T. & Kluemper, D. (2007). Evidence of predictive and incremental validity of personality factors, vocal attractiveness and the situational interview. *International Journal of Selection and Assessment*, *15*(1), 30–39. (Cit. on pp. 29, 30).

DeGroot, T. & Motowidlo, S. J. (1999). Why visual and vocal interview cues can affect interviewers' judgments and predict job performance. *Journal of Applied Psychology*, *84*(6), 986. (Cit. on pp. 29, 30).

Dellwo, V., Leemann, A. & Kolly, M.-J. (2015). The recognition of read and spontaneous speech in local vernacular: The case of Zurich German. *Journal of Phonetics*, *48*, 13–28. (Cit. on p. 88).

DePaulo, B. M., Blank, A. L., Swaim, G. W. & Hairfield, J. G. (1992, June). Expressiveness and expressive control. *Personality and Social Psychology Bulletin*, *18*(3), 276–285. (Cit. on p. 124).

DeSteno, D., Breazeal, C., Frank, R. H., Pizarro, D., Baumann, J., Dickens, L. & Lee, J. J. (2012, December). Detecting the trustworthiness of novel partners in economic exchange. *Psychological Science*, *23*(12), 1549–1556. (Cit. on p. 150).

Diekman, A. B. & Eagly, A. H. (2000). Stereotypes as dynamic constructs: Women and men of the past, present, and future. *Personality and Social Psychology Bulletin*, *26*(10), 1171–1188. (Cit. on p. 103).

Digman, J. M. (1990). Personality structure: Emergence of the five-factor model. *Annual Review of Psychology*, *41*(1), 417–440. (Cit. on pp. 27, 95).

Dion, K., Berscheid, E. & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285. (Cit. on pp. 28, 123).

Dixon, J. A., Mahoney, B. & Cocks, R. (2002, June). Accents of guilt?: Effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology*, *21*(2), 162–168. (Cit. on pp. 34, 59, 60, 72, 79, 92, 98, 99).

Dobrow, J. R. & Gidney, C. L. (1998). The good, the bad, and the foreign: The use of dialect in children's animated television. *The Annals of the American Academy of Political and Social Science*, 105–119. (Cit. on p. 33).

Doebeli, M., Hauert, C. & Killingback, T. (2004). The evolutionary origin of cooperators and defectors. *science*, *306*(5697), 859–862. (Cit. on pp. 10, 23, 24).

Donaldson, Z. R. & Young, L. J. (2008). Oxytocin, vasopressin, and the neurogenetics of sociality. *Science*, *322*(5903), 900–904. (Cit. on p. 25).

Dragojevic, M. & Giles, H. (2014). The reference frame effect: An intergroup perspective on language attitudes. *Human Communication Research*, *40*(1), 91–111. (Cit. on p. 33).

Drahota, A., Costall, A. & Reddy, V. (2008, April). The vocal communication of different kinds of smile. *Speech Communication*, *50*(4), 278–287. (Cit. on pp. 124, 125, 127).

Duffy, B. R. & Joue, G. (2004). I, robot being. In *Proceedings of the 8th Intelligent Autonomous Systems Conference (IAS8)* (pp. 120–127). Amsterdam, The Netherlands. (Cit. on p. 145).

Duffy, S. & Smith, J. (2014). Cognitive load in the multi-player prisoner's dilemma game: Are there brains in games? *Journal of Behavioral and Experimental Economics*, *51*, 47–56. (Cit. on p. 187).

Dunn, J. R. & Schweitzer, M. E. (2005). Feeling and believing: The influence of emotion on trust. *Journal of Personality and Social Psychology*, *88*(5), 736–748. (Cit. on p. 40).

Eagly, A. H., Wood, W. & Diekman, A. B. (2000). Social role theory of sex differences and similarities: A current appraisal. In T. Eckes & H. M. Trautner (Eds.), *The developmental social psychology of gender* (pp. 123–174). (Cit. on p. 103).

Eckel, C. C. & Wilson, R. K. (2003). Conditional trust: Sex, race and facial expressions in an experimental trust game. *Virginia Tech University Department of Economics Working Paper*. (Cit. on p. 112).

Eckel, C. C. & Wilson, R. K. (2004). Is trust a risky decision? *Journal of Economic Behavior & Organization*, *55*(4), 447–465. (Cit. on p. 55).

Eckel, C. C. & Wilson, R. K. (2005). Attractiveness and trust: Does beauty confound intuition? *Virginia Tech University Department of Economics Working Paper*. (Cit. on p. 112).

Eisenbarth, H. & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion*, *11*(4), 860–865. (Cit. on p. 122).

Ekman, P. (1992). Are there basic emotions? *Psychological Review*, *99*(3), 550–553. (Cit. on p. 122).

Ekman, P. (1999, January). Basic emotions. In P. Dalgleish & M. Power (Eds.), *Handbook of cognition and emotion* (Chap. 3, pp. 45–60). John Wiley & Sons, Ltd. (Cit. on p. 121).

Ekman, P., Davidson, R. J. & Friesen, W. V. (1990). The duchenne smile: Emotional expression and brain physiology: II. *Journal of Personality and Social Psychology*, *58*(2), 342–353. (Cit. on p. 123).

Ekman, P. & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology*, *17*(2), 124–129. (Cit. on p. 122).

Ekman, P. & Friesen, W. V. (1982). Felt, false, and miserable smiles. *Journal of Nonverbal Behavior*, *6*(4), 238–252. (Cit. on pp. 123, 140).

El Haddad, K., Cakmak, H., Moinet, A., Dupont, S. & Dutoit, T. (2015). An HMM approach for synthesizing amused speech with a controllable intensity of smile. In *Signal Processing and Information Technology (ISSPIT), 2015 IEEE International Symposium on* (pp. 7–11). IEEE. A. (Cit. on pp. 124–126).

El Haddad, K., Dupont, S., d'Alessandro, N. & Dutoit, T. (2015). An HMM-based speech-smile synthesis system: An approach for amusement synthesis. In *Proceedings of the 3rd International Workshop on Emotion Representation, Analysis and Synthesis in Continuous Time and Space (EmoSPACE'15)*. Ljubljana, Slovenia. (Cit. on p. 125).

Elfenbein, H. A. & Ambady, N. (2002). On the universality and cultural specificity of emotion recognition: A meta-analysis. *Psychological Bulletin*, *128*(2), 203–235. (Cit. on p. 122).

Elkins, A. C. & Derrick, D. C. (2013). The sound of trust: Voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation*, *22*(5), 897–913. (Cit. on pp. 12, 14, 38, 39, 77, 97, 99, 104, 123).

Émond, C. & Laforest, M. (2013). Prosodic correlates of smiled speech. In *Proceedings of Meetings on Acoustics (ICA2013)* (Vol. 19, *1*, p. 060220). 166th Meeting of the Acoustical Society of America. ASA. S. (Cit. on p. 124).

Engle-Warnick, J. & Slonim, R. L. (2006, January). Learning to trust in indefinitely repeated games. *Games and Economic Behavior*, *54*(1), 95–114. (Cit. on p. 14)

Epley, N., Waytz, A. & Cacioppo, J. T. (2007). On seeing human: A three-factor theory of anthropomorphism. *Psychological Review*, *114*(4), 864–886. (Cit. on p. 145).

Evans, B. G. & Iverson, P. (2004, January). Vowel normalization for accent: An investigation of best exemplar locations in northern and southern british english sentences. *The Journal of the Acoustical Society of America*, *115*(1), 352–361. (Cit. on p. 72).

Eyssel, F., Kuchenbrandt, D., Hegel, F. & de Ruiter, L. (2012, September). Activating elicited agent knowledge: How robot and user features shape the perception of social robots. In *2012 IEEE RO-MAN: The 21st IEEE international symposium on robot and human interactive communication* (pp. 851–857). IEEE. IEEE. (Cit. on p. 167).

Fagel, S. (2009). Effects of smiled speech on lips, larynx and acoustics. In *Proceedings of the Auditory Visual Speech Processing Conference (AVSP)* (pp. 18–21). N. (Cit. on pp. 124, 127).

Fehr, E. (2009a). On the economics and biology of trust. *Journal of the European Economic Association*, *7*(2-3), 235–266. (Cit. on pp. 22, 46, 55).

Fehr, E. (2009b). Social preferences and the brain. In P. W. Glimcher, C. F. Camerer, E. Fehr & R. A. Poldrack (Eds.), *Neuroeconomics: Decision making and the brain* (pp. 215–232). Amsterdam, The Netherlands: Elsevier. (Cit. on p. 26).

Fehr, E. & Gächter, S. (1998). Reciprocity and economics: The economic implications of homo reciprocans. *European Economic Review*, *42*(3), 845–859. (Cit. on p. 184).

Feinberg, D. R., DeBruine, L. M., Jones, B. C. & Perrett, D. I. (2008). The role of femininity and averageness of voice pitch in aesthetic judgments of women's voices. *Perception*, *37*(4), 615–623. (Cit. on pp. 30, 102, 115).

Feinberg, D. R., Jones, B. C., DeBruine, L. M., Moore, F. R., Smith, M. J. L., Cornwell, R. E., . . . Perrett, D. I. (2005). The voice and face of woman: One ornament that signals quality? *Evolution and Human Behavior*, *26*(5), 398–408. (Cit. on p. 102).

Feinberg, D. R., Jones, B. C., Little, A. C., Burt, D. M. & Perrett, D. I. (2005). Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*(3), 561–568. (Cit. on pp. 102, 114).

Fetchenhauer, D. & Dunning, D. (2012). Betrayal aversion versus principled trustfulness—how to explain risk avoidance and risky choices in trust games. *Journal of Economic Behavior & Organization*, *81*(2), 534–541. (Cit. on p. 55).

Fiedler, K. & Walka, I. (1993). Training lie detectors to use nonverbal cues instead of global heuristics. *Human Communication Research*, *20*(2), 199–223. (Cit. on p. 76).

Fitch, W. T. (1997). Vocal tract length and formant frequency dispersion correlate with body size in rhesus macaques. *The Journal of the Acoustical Society of America*, *102*(2), 1213–1222. (Cit. on pp. 102, 114).

Fitch, W. T. (1999). Acoustic exaggeration of size in birds via tracheal elongation: Comparative and theoretical analyses. *Journal of Zoology*, *248*(1), 31–48. (Cit. on p. 114).

Fitch, W. T. & Giedd, J. (1999). Morphology and development of the human vocal tract: A study using magnetic resonance imaging. *The Journal of the Acoustical Society of America*, *106*(3), 1511–1522. (Cit. on p. 114).

Floccia, C., Goslin, J., Girard, F. & Konopczynski, G. (2006). Does a regional accent perturb speech processing? *Journal of Experimental Psychology: Human Perception and Performance*, *32*(5), 1276–1293. (Cit. on p. 72).

Folstad, I. & Karter, A. J. (1992). Parasites, bright males, and the immunocompetence handicap. *The American Naturalist*, *139*(3), 603–622. (Cit. on p. 102).

Fraccaro, P. J., O'Connor, J. J. M., Re, D. E., Jones, B. C., DeBruine, L. M. & Feinberg, D. R. (2013). Faking it: Deliberately altered voice pitch and vocal attractiveness. *Animal Behaviour*, *85*(1), 127–136. (Cit. on pp. 114, 115, 117).

Francis, K. B., Howard, C., Howard, I. S., Gummerum, M., Ganis, G., Anderson, G. & Terbeck, S. (2016). Virtual morality: Transitioning from moral judgment to moral action? *PloS one*, *11*(10). (Cit. on p. 16).

Franck, E., De Raedt, R. & De Houwer, J. (2007). Implicit but not explicit self-esteem predicts future depressive symptomatology. *Behaviour Research and Therapy*, *45*(10), 2448–2455. (Cit. on p. 11).

Frank, R. H. (1988). *Passions within reason: The strategic role of the emotions.* WW Norton & Co. (Cit. on p. 124).

Frank, R. H., Gilovich, T. & Regan, D. T. (1993). The evolution of one-shot cooperation: An experiment. *Ethology and sociobiology*, *14*(4), 247–256. (Cit. on p. 183).

Frick, R. W. (1985). Communicating emotion: The role of prosodic features. *Psychological Bulletin*, *97*(3), 412–429. (Cit. on p. 40).

Frumkin, L. (2007). Influences of accent and ethnic background on perceptions of eyewitness testimony. *Psychology, Crime and Law*, *13*, 317–331. (Cit. on pp. 69, 92).

Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C. & Giles, H. (2012, October). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*(1), 120–133. (Cit. on pp. 32, 34, 59, 60, 68, 69, 72, 79, 92, 98, 99).

Fussell, S. R., Kiesler, S., Setlock, L. D. & Yew, V. (2008). How people anthropomorphize robots. In *Proceedings of the 3rd ACM/IEEE International Conference on Human-Robot Interaction* (pp. 145–152). ACM. (Cit. on p. 145).

Gambetta, D. (1988). Can we trust trust? In D. Gambetta (Ed.), *Trust: Making and breaking cooperative relations* (pp. 213–237). Basil Blackwell. (Cit. on pp. 13, 22, 54).

Garellek, M., Keating, P., Esposito, C. M. & Kreiman, J. (2013). Voice quality and tone identification in White Hmong. *Journal of the Acoustical Society of America*, *133*, 1078–1089. (Cit. on p. 81).

Garrett, P., Coupland, N. & Williams, A. (1999). Evaluating dialect in discourse: Teachers' and teenagers' responses to young English speakers in Wales. *Language in Society*, *28*(03), 321–354. (Cit. on p. 34).

Garrett, P., Williams, A. & Evans, B. (2005). Attitudinal data from New Zealand, Australia, the USA and UK about each other's Englishes: Recent changes or consequences of methodologies? *Multilingua*, *24*(3), 211–235. (Cit. on p. 34).

Gay, T. (1981). Mechanisms in the control of speech rate. *Phonetica*, *38*(1-3), 148–158. (Cit. on p. 104).

Giammanco, M., Tabacchi, G., Giammanco, S., Di Majo, D. & La Guardia, M. (2005). Testosterone and aggressiveness. *Medical Science Monitor*, *11*(4), RA136–RA145. (Cit. on p. 102).

Giles, H. (1970). Evaluative reactions to accents. *Educational review*, *22*(3), 211–227. (Cit. on pp. 34, 35, 72, 98, 99).

Giles, H. (1971). Patterns of evaluation to RP, South Welsh and Somerset accented speech. *British Journal of Social and Clinical Psychology*, *10*(3), 280–281. (Cit. on pp. 33, 93, 94, 175).

Giles, H. & Billings, A. C. (2004). Assessing language attitudes: Speaker evaluation studies. In A. Davis & E. Elder (Eds.), *The handbook of applied linguistics* (pp. 187–209). Blackwell Publishing Ltd. (Cit. on p. 34).

Giles, H. & Powesland, P. (1997). Accommodation theory. In N. Coupland & A. Jaworski (Eds.), *Sociolinguistics* (pp. 232–239). London: Macmillan Education UK. (Cit. on pp. 14, 95, 104, 117).

Giles, H., Williams, A., Mackie, D. M. & Rosselli, F. (1995). Reactions to Anglo-and Hispanic-American-accented speakers: Affect, identity, persuasion, and the English-only controversy. *Language & Communication*, *15*(2), 107–120. (Cit. on pp. 60, 72).

Glaeser, E. L., Laibson, D. I., Scheinkman, J. A. & Soutter, C. L. (2000). Measuring trust. *Quarterly Journal of Economics*, 811–846. (Cit. on pp. 11, 15, 16, 26).

Gobl, C. & Ní Chasaide, A. (2003). The role of voice quality in communicating emotion, mood and attitude. *Speech Communication*, *40*, 189–212. (Cit. on p. 165).

Goggin, J. P., Thompson, C. P., Strube, G. & Simental, L. R. (1991). The role of language familiarity in voice identification. *Memory & Cognition*, *19*(5), 448–458. (Cit. on p. 34).

Goldberg, L. R. (1990). An alternative" description of personality": The big-five factor structure. *Journal of Personality and Social Psychology*, *59*(6), 1216. (Cit. on p. 95).

Goldberg, L. R., Johnson, J. A., Eber, H. W., Hogan, R., Ashton, M. C., Cloninger, C. R. & Gough, H. G. (2006, February). The international personality item pool and the future of public-domain personality measures. *Journal of Research in personality*, *40*(1), 84–96. (Cit. on p. 156).

Gooskens, C. (2005). How well can Norwegians identify their dialects? *Nordic Journal of Linguistics*, *28*(01), 37–60. (Cit. on p. 35).

Grammer, K., Fink, B., Møller, A. P. & Thornhill, R. (2003). Darwinian aesthetics: Sexual selection and the biology of beauty. *Biological Reviews*, *78*(3), 385–407. (Cit. on p. 30).

Graziano, W. G. & Eisenberg, N. (1997). Agreeableness: A dimension of personality. In R. Hogan, J. Johnson & S. Briggs (Eds.), *Handbook of personality psychology* (pp. 795–824). Academic Press. (Cit. on p. 27).

Graziano, W. G. & Tobin, R. M. (2002). Agreeableness: Dimension of personality or social desirability artifact? *Journal of personality*, *70*(5), 695–728. (Cit. on p. 27).

Greenwald, A. G. (1990). What cognitive representations underlie social attitudes? *Bulletin of the Psychonomic Society*, *28*(3), 254–260. On Implicit vs explicit tasks. (Cit. on pp. 11, 45).

Greenwald, A. G. & Banaji, M. R. (1995). Implicit social cognition: Attitudes, self-esteem, and stereotypes. *Psychological Review*, *102*(1), 4. (Cit. on pp. 11, 45).

Greenwald, A. G., McGhee, D. E. & Schwartz, J. L. K. (1998). Measuring individual differences in implicit cognition: The implicit association test. *Journal of Personality and Social Psychology*, *74*(6), 1464. (Cit. on p. 11).

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. L. Morgan (Eds.), *Syntax and semantics: Speech acts* (Vol. 3, pp. 41–58). New York: Academic Press. (Cit. on p. 10).

Groom, V. & Nass, C. I. (2007). Can robots be teammates? Benchmarks in human-robot teams. *Interaction Studies*, *8*(3), 483–500. (Cit. on p. 146).

Gussenhoven, C. (2002, April 13). Intonation and interpretation: Phonetics and phonology. In *Proceedings of the First International Conference on Speech Prosody*. Aix-en-Provence, France. (Cit. on pp. 95, 104, 116).

Gussenhoven, C. (2016). Foundations of intonational meaning: Anatomical and physiological factors. *Topics in Cognitive Science*, *8*(2), 425–434. (Cit. on pp. 40, 77).

Ham, J., Bokhorst, R., Cuijpers, R., van der Pol, D. & Cabibihan, J.-J. (2011). Making robots persuasive: The influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In *Proceedings of the 3rd International Conference on Social Robotics (ICSR)* (pp. 71–83). Springer. Amsterdam, The Netherlands. (Cit. on p. 148).

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y. C., De Visser, E. J. & Parasuraman, R. (2011). 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. (Cit. on pp. 146, 150, 166).

Harcourt, A. H. (1991). Help, cooperation and trust in animals. In R. A. Hinde & J. Groebel (Eds.), *Cooperation and prosocial behaviour* (pp. 15–26). Cambridge University Press. (Cit. on p. 23).

Haring, K. S., Matsumoto, Y. & Watanabe, K. (2013). How do people perceive and trust a lifelike robot. In *Lecture Notes in Engineering and Computer Science: Proceedings of The World Congress on Engineering and Computer Science 2013* (Vol. 1, pp. 23–25). (Cit. on pp. 11, 149, 150).

Hatfield, E., Cacioppo, J. T. & Rapson, R. L. (1994). *Emotional contagion*. Cambridge University Press. (Cit. on p. 140).

Hauser, M. D. (1993). The evolution of nonhuman primate vocalizations: Effects of phylogeny, body weight, and social context. *The American Naturalist*, *142*(3), 528–542. (Cit. on p. 114).

Henton, C. G. & Bladon, R. A. W. (1985). Breathiness in normal female speech: Inefficiency versus desirability. *Language & Communication*, *5*(3), 221–227. (Cit. on p. 30).

Hewlett, N. & Rendall, M. (1998). Rural versus urban accent as an influence on the rate of speech. *Journal of the International Phonetic Association*, *28*(1-2), 63–71. (Cit. on p. 103).

Higgins, E. T. (2000). Making a good decision: Value from fit. *American Psychologist*, *55*(11), 1217. (Cit. on pp. 96, 117).

Hinds, P. J., Roberts, T. L. & Jones, H. (2004). Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, *19*(1), 151–181. (Cit. on p. 170).

Hiraga, Y. (2005). British attitudes towards six varieties of English in the USA and Britain. *World Englishes*, *24*(3), 289–308. (Cit. on pp. 34, 79, 92).

Hirose, K. & Kawanami, H. (2002). Temporal rate change of dialogue speech in prosodic units as compared to read speech. *Speech Communication*, *36*(1), 97–111. (Cit. on p. 88).

Hirschberg, J. (2002, April 13). The Pragmatics of Intonational Meaning. In *Proceedings of the First International Conference on Speech Prosody*. Aix-en-Provence, France. (Cit. on p. 95).

Hodges-Simeon, C. R., Gaulin, S. J. C. & Puts, D. A. (2010). Different vocal parameters predict perceptions of dominance and attractiveness. *Human Nature*, *21*(4), 406–427. (Cit. on pp. 30, 102, 115).

Hodges-Simeon, C. R., Gurven, M., Puts, D. A. & Gaulin, S. J. C. (2014). Vocal fundamental and formant frequencies are honest signals of threat potential in peripubertal males. *Behavioral Ecology*, aru081. (Cit. on p. 114).

Hoffman, E., McCabe, K. & Smith, V. L. (1996). Social distance and other-regarding behavior in dictator games. *The American Economic Review*, *86*(3), 653–660. (Cit. on p. 184).

Hofmann, W., Gawronski, B., Gschwendner, T., Le, H. & Schmitt, M. (2005). A meta-analysis on the correlation between the Implicit Association Test and explicit self-report measures. *Personality and Social Psychology Bulletin*, *31*(10), 1369–1385. (Cit. on p. 45).

Houser, D., Schunk, D. & Winter, J. (2010). Distinguishing trust from risk: An anatomy of the investment game. *Journal of Economic Behavior & Organization*, *74*(1), 72–81. (Cit. on p. 55).

Hughes, S. M., Dispenza, F. & Gallup, G. G. (2004). Ratings of voice attractiveness predict sexual behavior and body configuration. *Evolution and Human Behavior*, *25*(5), 295–304. (Cit. on p. 30).

Hughes, S. M., Pastizzo, M. J. & Gallup, G. G. (2008). The sound of symmetry revisited: Subjective and objective analyses of voice. *Journal of Nonverbal Behavior*, *32*(2), 93–108. (Cit. on pp. 38, 76).

Hwang, J., Park, T. & Hwang, W. (2013). The effects of overall robot shape on the emotions invoked in users and the perceived personalities of robot. *Applied ergonomics*, *44*(3), 459–471. (Cit. on p. 42).

Imhof, M. (2010). Listening to voices and judging people. *The International Journal of Listening*, *24*, 19–33. (Cit. on pp. 39, 76, 85, 95, 99, 103).

Izard, C. E. (1969). The emotions and emotion constructs in personality and culture research. In R. B. Cattell & R. M. Dreger (Eds.), *Handbook of modern personality theory* (pp. 496–510). Wiley New York. (Cit. on p. 122).

Jacewicz, E., Fox, R. A., O'Neill, C. & Salmons, J. (2009, July). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(02), 233. (Cit. on pp. 37, 81, 103).

Jacewicz, E., Fox, R. A. & Wei, L. (2010). Between-speaker and within-speaker variation in speech tempo of American English. *The Journal of the Acoustical Society of America*, *128*(2), 839–850. (Cit. on p. 88).

Jackson, L. A., Hunter, J. E. & Hodge, C. N. (1995). Physical attractiveness and intellectual competence: A meta-analytic review. *Social Psychology Quterly*, 108–122. (Cit. on p. 28).

Jiang, X. & Pell, M. D. (2017). The sound of confidence and doubt. *Speech Communication*, *88*, 106–126. (Cit. on pp. 37, 76, 116).

Johnson, K. (2002). *Acoustic and Auditory Phonetics* (2nd). Oxford: Blackwell. (Cit. on p. 81).

Jones, B. C., Feinberg, D. R., DeBruine, L. M., Little, A. C. & Vukovic, J. (2010). A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Animal Behaviour*, *79*(1), 57–62. (Cit. on pp. 30, 102, 103, 113).

Jones, D. N. & Paulhus, D. L. (2009). Machiavellianism. In M. R. Leary & R. H. Hoyle (Eds.), *Handbook of individual differences in social behavior* (pp. 93–108). New York, NY, US: Guilford Press. (Cit. on p. 26).

Jones, G. R. & George, J. M. (1998). The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of Management Review*, *23*(3), 531–546. (Cit. on pp. 13, 25, 26, 42).

Jonker, C. M. & Treur, J. (1999). Formal analysis of models for the dynamics of trust based on experiences. In *Proceedings of the 9th European Workshop on Modelling Autonomous Agents in a Multi-Agent World* (pp. 221–231). Springer. Valencia, Spain. (Cit. on p. 13).

Jordan, J. J., Hoffman, M., Nowak, M. A. & Rand, D. G. (2016). Uncalculating cooperation is used to signal trustworthiness. *Proceedings of the National Academy of Sciences*, *113*(31), 8658–8663. (Cit. on p. 49).

Joshi, J., Gunes, H. & Goecke, R. (2014). Automatic prediction of perceived traits using visual cues under varied situational context. In *Proceedings of the 22nd International Conference on Pattern Recognition (ICPR)* (pp. 2855–2860). IEEE. Stockholm, Sweden. (Cit. on p. 13).

Kahneman, D., Knetsch, J. L. & Thaler, R. H. (1986). Fairness and the assumptions of economics. *Journal of business*, S285–S300. (Cit. on p. 184).

Kappas, A., Hess, U. & Scherer, K. R. (1991). Voice and emotion. In R. S. Feldman & B. Rimé (Eds.), *Fundamentals of nonverbal behavior* (pp. 200–238). Studies in Emotion & Social Interaction. Cambridge University Press. (Cit. on p. 40).

Karlan, D. S. (2005, November). Using experimental economics to measure social capital and predict financial decisions. *The American Economic Review*, *95*(5), 1688–1699. (Cit. on p. 55).

King-Casas, B., Tomlin, D., Anen, C., Camerer, C. F., Quartz, S. R. & Montague, P. R. (2005). Getting to know you: Reputation and trust in a two-person economic exchange. *Science*, *308*(5718), 78–83. (Cit. on pp. 13, 49).

Kinzler, K. D., Corriveau, K. H. & Harris, P. L. (2011). Children's selective trust in native-accented speakers. *Developmental Science*, *14*(1), 106–111. (Cit. on p. 33).

Kinzler, K. D. & DeJesus, J. M. (2013). Northern= smart and Southern= nice: The development of accent attitudes in the United States. *The Quarterly Journal of Experimental Psychology*, *66*(6), 1146–1158. (Cit. on p. 33).

Kinzler, K. D., Shutts, K., DeJesus, J. M. & Spelke, E. S. (2009). Accent trumps race in guiding children's social preferences. *Social Cognition*, *27*(4), 623. (Cit. on p. 32).

Kirchhübel, C. & Howard, D. M. (2013). Detecting suspicious behaviour using speech: Acoustic correlates of deceptive speech - an exploratory investigation. *Applied Ergonomics*, *44*, 694–702. (Cit. on pp. 38, 75).

Kitayama, S., Mesquita, B. & Karasawa, M. (2006). Cultural affordances and emotional experience: Socially engaging and disengaging emotions in japan and the united states. *Journal of Personality and Social Psychology*, *91*(5), 890–903. (Cit. on p. 122).

Klatt, D. H. & Klatt, L. C. (1990, February). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *The Journal of the Acoustical Society of America*, *87*(2), 820–857. (Cit. on pp. 39, 76).

Klinnert, M. D., Campos, J. J., Sorce, J. F., Emde, R. N. & Svejda, M. (1983). Emotions as behavior regulators: Social referencing in infants. In R. Plutchik & H. Kellerman (Eds.), *Emotion theory, research, and experience* (Vol. 2, pp. 57–68). New York: Academic Press. (Cit. on p. 122).

Klofstad, C. A. (2015). Candidate voice pitch influences election outcomes. *Political Psychology*. (Cit. on pp. 38, 77).

Klofstad, C. A., Anderson, R. C. & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society of London B: Biological Sciences*, *279*(1738), 2698–2704. (Cit. on pp. 102, 113).

Knack, S. & Keefer, P. (1997, November). Does social capital have an economic payoff? a cross-country investigation. *The Quarterly journal of economics*, *112*(4), 1251–1288. (Cit. on p. 23).

Knight, C. (1998). Ritual/speech coevolution: A solution to the problem of deception. In J. R. Hurford, M. Studdert-Kennedy & C. Knight (Eds.), *Evolution of Language: Social and Cognitive Bases* (pp. 68–91). Cambridge University Press. (Cit. on p. 27).

Knowles, K. K. & Little, A. C. (2016). Vocal fundamental and formant frequencies affect perceptions of speaker cooperativeness. *The Quarterly Journal of Experimental Psychology*, *69*(9), 1657–1675. (Cit. on pp. 39, 76, 103, 113).

Ko, S. J., Judd, C. M. & Blair, I. V. (2006). What the voice reveals: Within-and between-category stereotyping on the basis of voice. *Personality and Social Psychology Bulletin*, *32*(6), 806–819. (Cit. on p. 113).

Koay, K. L., Syrdal, D. S., Walters, M. L. & Dautenhahn, K. (2007). Living with robots: Investigating the habituation effect in participants' preferences during a longitudinal human-robot interaction study. In *RO-MAN 2007 - the 16th IEEE international symposium on robot and human interactive communication* (pp. 564–569). IEEE. IEEE. (Cit. on pp. 149, 166).

Kohler, C. G., Turner, T. H., Gur, R. E. & Gur, R. C. (2004, April). Recognition of facial emotions in neuropsychiatric disorders. *CNS Spectrums*, *9*(04), 267–274. (Cit. on p. 121).

Kohler, K. J. (2008). 'speech-smile','speech-laugh','laughter'and their sequencing in dialogic interaction. *Phonetica*, *65*(1-2), 1–18. (Cit. on pp. 123, 124, 127, 140).

Koscik, T. R. & Tranel, D. (2011). The human amygdala is necessary for developing and expressing normal interpersonal trust. *Neuropsychologia*, *49*(4), 602–611. (Cit. on p. 25).

Kosfeld, M. (2007, July). Trust in the brain. neurobiological determinants of human social behaviour. *EMBO reports*, *8*(1S), S44–S47. (Cit. on p. 25).

Kosfeld, M., Heinrichs, M., Zak, P. J., Fischbacher, U. & Fehr, E. (2005, June). Oxytocin increases trust in humans. *Nature*, *435*(7042), 673–676. (Cit. on p. 55).

Krach, S., Hegel, F., Wrede, B., Sagerer, G., Binkofski, F. & Kircher, T. (2008). Can machines think? Interaction and perspective taking with robots investigated via fMRI. *PloS ONE*, *3*(7), e2597. (Cit. on p. 145).

Kriengwatana, B., Terry, J., Chládková, K. & Escudero, P. (2016). Speaker and accent variation are handled differently: Evidence in native and non-native listeners. *PloS One*, *11*(6), e0156870. (Cit. on p. 36).

Kristiansen, C. M., Eiser, J. R., Harding, C. M. & O'Looney, B. A. (1983). Inferences of a speaker's personality and health behaviour from accented speech. *Journal of the Institute of Health Education*, *21*(4), 113–120. (Cit. on pp. 34, 94).

Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L. & Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion*, *7*(4), 730–735. (Cit. on pp. 49, 123, 139).

La Porta, R., Lopez-De-Silane, F., Shleifer, A. & Vishny, R. W. (1997). Trust in large organizations. *American Economic Review*, *87*(2), 333–338. (Cit. on p. 23).

Laan, G. P. M. (1997). The contribution of intonation, segmental durations, and spectral features to the perception of a spontaneous and a read speaking style. *Speech Communication*, *22*(1), 43–65. (Cit. on p. 88).

Ladegaard, H. J. (1998). National stereotypes and language attitudes: The perception of British, American and Australian language and culture in Denmark. *Language & Communication*, *18*(4), 251–274. (Cit. on pp. 34, 35, 59).

Lagenspetz, O. (1992). Legitimacy and trust. *Philosophical Investigations*, *15*(1), 1–21. (Cit. on p. 22).

Lambert, W. E., Hodgson, R. C., Gardner, R. C. & Fillenbaum, S. (1960). Evaluational reactions to spoken languages. *The Journal of Abnormal and Social Psychology*, *60*(1), 44. (Cit. on p. 36).

Langlois, J. H., Kalakanis, L., Rubenstein, A. J., Larson, A., Hallam, M. & Smoot, M. (2000). Maxims or myths of beauty? a meta-analytic and theoretical review. *Psychological bulletin*, *126*(3), 390. (Cit. on p. 28).

Lanzetta, J. T. & Englis, B. G. (1989). Expectations of cooperation and competition and their effects on observers' vicarious emotional responses. *Journal of Personality and Social Psychology*, *56*(4), 543–554. (Cit. on p. 125).

Large, D. R., Clark, L., Quandt, A., Burnett, G. E. & 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. (Cit. on pp. 18, 41, 42).

Lasarcyk, E. & Trouvain, J. (2008). Spread lips+ raised larynx+ higher f0= smiled speech?-an articulatory synthesis approach. In *Proceedings of the 8th International Seminar on Speech Production (ISSP'08)* (pp. 43–48). Strasbourg, France. (Cit. on pp. 125, 127).

Lau, S. (1982, June). The effect of smiling on person perception. *The Journal of Social Psychology*, *117*(1), 63–67. (Cit. on pp. 123, 139).

Laver, J. D. M. (1968). Voice quality and indexical information. *British Journal of Disorders of Communication*, *3*(1), 43–54. (Cit. on pp. 30, 40, 76).

Lee, K. M., Peng, W., Jin, S.-A. & Yan, C. (2006). 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. (Cit. on pp. 41, 147, 170).

Lemaignan, S., Fink, J. & Dillenbourg, P. (2014). The dynamics of anthropomorphism in robotics. In *Proceedings of the 9th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 226–227). ACM. (Cit. on p. 145).

Lev-Ari, S. & Keysar, B. (2010, November). Why don't we believe non-native speakers? the influence of accent on credibility. *Journal of Experimental Social Psychology*, *46*(6), 1093–1096. (Cit. on pp. 10, 33, 59, 69, 92).

Lewicki, R. J., Tomlinson, E. C. & Gillespie, N. (2006). Models of interpersonal trust development: Theoretical approaches, empirical evidence, and future directions. *Journal of Management*, *32*(6), 991–1022. (Cit. on pp. 13, 15).

Lippi-Green, R. (1997). *English with an accent: Language, ideology, and discrimination in the United States*. Psychology Press. (Cit. on pp. 31, 32).

Lipson, S. F. & Ellison, P. T. (1996). Comparison of salivary steroid profiles in naturally occurring conception and non-conception cycles. *Human Reproduction*, *11*(10), 2090–2096. (Cit. on p. 102).

Liu, X. & Xu, Y. (2011). What makes a female voice attractive. In *Proceedings of the 17th International Congress of Phonetic Sciences*. Hong Kong. (Cit. on p. 30).

Lount, J., Robert B. (2010). The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of Personality and Social Psychology*, *98*(3), 420–433. (Cit. on p. 122).

Luhmann, N. (1979). *Trust and power*. John Wiley & Sons Inc. (Cit. on pp. 22, 23).

Lumsden, J., Edwards, E. A., Lawrence, N. S., Coyle, D. & Munafò, M. R. (2016). Gamification of cognitive assessment and cognitive training: A systematic review of applications and efficacy. *JMIR serious games*, *4*(2). (Cit. on p. 16).

Mack, S. & Munson, B. (2012). The influence of/s/quality on ratings of men's sexual orientation: Explicit and implicit measures of the 'gay lisp' stereotype. *Journal of Phonetics*, *40*(1), 198–212. (Cit. on p. 36).

Macy, M. W. & Skvoretz, J. (1998). The evolution of trust and cooperation between strangers: A computational model. *American Sociological Review*, 638–660. (Cit. on p. 166).

Magnée, M. J. C. M., Stekelenburg, J. J., Kemner, C. & de Gelder, B. (2007, March). Similar facial electromyographic responses to faces, voices, and body expressions. *NeuroReport*, *18*(4), 369–372. (Cit. on p. 140).

Marno, H., Guellai, B., Vidal, Y., Franzoi, J., Nespor, M. & Mehler, J. (2016). Infants' selectively pay attention to the information they receive from a native speaker of their language. *Frontiers in Psychology*, *7*, 1150. (Cit. on p. 33)

Mason, M. F., Tatkow, E. P. & Macrae, C. N. (2005). The look of love. gaze shifts and person perception. *Psychological Science*, *16*(3), 236–239. (Cit. on p. 148).

Matsumoto, D. (1993, June). Ethnic differences in affect intensity, emotion judgments, display rule attitudes, and self-reported emotional expression in an american sample. *Motivation and Emotion*, *17*(2), 107–123. (Cit. on p. 122).

Mayer, R. C., Davis, J. H. & Schoorman, F. D. (1995, July). An integrative model of organizational trust. *The Academy of Management Review*, *20*(3), 709–734. (Cit. on pp. 21, 22).

Mazzella, R. & Feingold, A. (1994). The effects of physical attractiveness, race, socioeconomic status, and gender of defendants and victims on judgments of mock jurors: A meta-analysis. *Journal of Applied Social Psychology*, *24*(15), 1315–1338. (Cit. on p. 28).

McAleer, P., Todorov, A. & Belin, P. (2014). How do you say Hello? Personality impressions from brief novel voices. *PLoS ONE*, *9*(3), e90779. (Cit. on pp. 10, 14, 15, 28, 37, 59, 104, 115).

McAllister, D. J. (1995). Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of Management Journal*, *38*(1), 24–59. (Cit. on pp. 102, 113).

McCabe, K., Houser, D., Ryan, L., Smith, V. & Trouard, T. (2001). A functional imaging study of cooperation in two-person reciprocal exchange. *Proceedings of the National Academy of Sciences*, *98*(20), 11832–11835. (Cit. on p. 55).

McCrae, R. R. (2009). The five-factor model of personality. In P. Corr & G. Mathhews (Eds.), *Handbook of Personality Psychology* (pp. 148–161). (Cit. on pp. 27, 39, 76, 113).

McCroskey, J. C. & Young, T. J. (1981). Ethos and credibility: The construct and its measurement after three decades. *Communication Studies*, *32*(1), 24–34. (Cit. on p. 156).

Mehu, M. & Dunbar, R. I. (2008). Relationship between smiling and laughter in humans (homo sapiens): Testing the power asymmetry hypothesis. *Folia Primatologica*, *79*(5), 269–280. (Cit. on p. 122).

Messinger, D. S., Fogel, A. & Dickson, K. L. (2001). All smiles are positive, but some smiles are more positive than others. *Developmental Psychology*, *37*(5), 642–653. (Cit. on p. 123).

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech* (pp. 39–74). Lawrence Erlbaum Associates. (Cit. on p. 104).

Miller, N., Maruyama, G., Beaber, R. J. & Valone, K. (1976). Speed of speech and persuasion. *Journal of Personality and Social Psychology*, *34*(4), 615. (Cit. on pp. 10, 37, 76, 104).

Minato, T., Shimada, M., Ishiguro, H. & Itakura, S. (2004, May 20). Development of an android robot for studying human-robot interaction. In *Innovations in applied artificial intelligence* (pp. 424–434). Springer Berlin Heidelberg. (Cit. on p. 146).

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. (Cit. on p. 147).

Montano, K. J., Tigue, C. C., Isenstein, S. G. E., Barclay, P. & Feinberg, D. R. (2017). Men's voice pitch influences women's trusting behavior. *Evolution and Human Behavior*, *38*(3), 293–297. (Cit. on pp. 49, 101, 113).

Moon, C., Cooper, R. P. & Fifer, W. P. (1993). Two-day-olds prefer their native language. *Infant behavior and development*, *16*(4), 495–500. (Cit. on p. 33).

Moore, C. & Dunham, P. (2014). *Joint attention: Its origins and role in development*. Psychology Press. (Cit. on p. 148).

Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy*, 7(4), 33-35. (Translated by Karl F. MacDorman and Takashi Minato in 2005 within Appendix B for the paper Androids as an Experimental Apparatus: Why is there an uncanny and can we exploit it? In *Proceedings of the CogSci-2005 Workshop: Toward Social Mechanisms of Android Science* (pp. 106–118). (Cit. on p. 146).

Moulines, E. & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*(5-6), 453–467. (Cit. on p. 106).

Mueser, K. T., Grau, B. W., Sussman, S. & Rosen, A. J. (1984). You're only as pretty as you feel: Facial expression as a determinant of physical attractiveness. *Journal of Personality and Social Psychology*, *46*(2), 469–478. (Cit. on p. 123).

Munro, M. J., Derwing, T. M. & Burgess, C. S. (2010). Detection of nonnative speaker status from content-masked speech. *Speech Communication*, *52*(7), 626–637. Reference for reversed utterances in order to mask speech content. (Cit. on p. 118).

Mutlu, B., Forlizzi, J. & Hodgins, J. (2006). A storytelling robot: Modeling and evaluation of human-like gaze behavior. In *Proceedings of the 6th IEEE-RAS International Conference on Humanoid Robots* (pp. 518–523). IEEE. Genova, Italy. (Cit. on p. 149).

Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H. & Hagita, N. (2009). Footing in human-robot conversations: How robots might shape participant roles using gaze cues. In *Proceedings of the 4th ACM/IEEE International Conference on Human–Robot Interaction* (pp. 61–68). ACM. San Diego, CA, USA. (Cit. on p. 148).

Nass, C. I. & Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship.* Cambridge, MA: MIT Press. (Cit. on pp. 39, 77, 146).

Nass, C. I. & Lee, K. M. (2001). Does computer-synthesized speech manifest personality? Experimental tests of recognition, similarity-attraction, and consistency-attraction. *Journal of Experimental Psychology: Applied*, *7*(3), 171–181. (Cit. on pp. 42, 146, 147, 166).

Nass, C. I., Moon, Y., Fogg, B. J., Reeves, B. & Dryer, C. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*(2), 223–239. (Cit. on pp. 17, 41, 147).

Nass, C. I., Steuer, J. & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 72–78). ACM. Boston, MA, USA. (Cit. on pp. 41, 45, 146, 166, 183).

Nazzi, T., Bertoncini, J. & Mehler, J. (1998). Language discrimination by newborns: Toward an understanding of the role of rhythm. *Journal of Experimental Psychology: Human perception and performance*, *24*(3), 756. (Cit. on p. 33).

Nettle, D. & Liddle, B. (2008). Agreeableness is related to social-cognitive, but not social-perceptual, theory of mind. *European Journal of Personality*, *22*(4), 323–335. (Cit. on pp. 27, 102, 113, 115).

Niebuhr, O., Brem, A., Novák-Tót, E. & Voße, J. (2016). 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*. B. (Cit. on pp. 37, 76, 95, 115, 116).

Niedzielski, N. A. & Preston, D. R. (2000). *Folk linguistics*. Trends in Linguistics: Studies and monographs. Walter de Gruyter. (Cit. on p. 31).

Nisbett, R. E. & Wilson, T. D. (1977). The halo effect: Evidence for unconscious alteration of judgments. *Journal of personality and social psychology*, *35*(4), 250. (Cit. on pp. 28, 72).

O'Connor, J. J. M. & Barclay, P. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*. (Cit. on pp. 30, 31, 38, 49, 75, 101, 102, 113–115).

O'Connor, J. J. M., Fraccaro, P. J., Pisanski, K., Tigue, C. C., O'Donnell, T. J. & Feinberg, D. R. (2014). Social dialect and men's voice pitch influence women's mate preferences. *Evolution and Human Behavior*, *35*(5), 368–375. (Cit. on pp. 102, 113).

Ohala, J. J. (1980). The acoustic origin of the smile. *Journal of the Acoustical Society of America*, *68*. (Cit. on pp. 95, 114, 122).

Ohala, J. J. (1983). Cross-language use of pitch: An ethological view. *Phonetica*, *40*, 1–18. (Cit. on pp. 95, 102, 114).

Omdal, H. (1995). Attitudes toward spoken and written Norwegian. *International Journal of the Sociology of Language*, *115*(1), 85–108. (Cit. on p. 34).

Oosterhof, N. N. & Todorov, A. (2009). Shared perceptual basis of emotional expressions and trustworthiness impressions from faces. *Emotion*, *9*(1), 128. (Cit. on p. 40).

Oswald, F. L., Mitchell, G., Blanton, H., Jaccard, J. & Tetlock, P. E. (2013). Predicting ethnic and racial discrimination: A meta-analysis of iat criterion studies. *Journal of personality and social psychology*, *105*(2), 171. (Cit. on pp. 11, 12).

Otta, E., Lira, B. B. P., Delevati, N. M., Cesar, O. P. & Pires, C. S. G. (1994, May). The effect of smiling and of head tilting on person perception. *The Journal of psychology*, *128*(3), 323–331. (Cit. on pp. 123, 139).

Packer, C. (1977, February). Reciprocal altruism in papio anubis. *Nature*, *265*(5593), 441–443. (Cit. on p. 23).

Patterson, D. & Ladd, D. R. (1999, August 7). Pitch Range Modelling: Linguistic Dimensions of Variation. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 1169–1172). S. (Cit. on p. 81).

Patzer, G. L. (1983). Source credibility as a function of communicator physical attractiveness. *Journal of business research*, *11*(2), 229–241. (Cit. on p. 28).

Penton-Voak, I. S., Pound, N., Little, A. C. & Perrett, D. I. (2006, October). Personality judgments from natural and composite facial images: More evidence for a "kernel of truth" in social perception. *Social Cognition*, *24*(5), 607–640. (Cit. on pp. 123, 139).

Pervin, L. A. (2001). *Personality: Theory and Research*. New York: Wiley. (Cit. on pp. 27, 39, 76, 113).

Pettigrew, T. F. & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. *Journal of Personality and Social Psychology*, *90*(5), 751. (Cit. on pp. 14, 104).

Piantadosi, S. T., Tily, H. & Gibson, E. (2011). Word lengths are optimized for efficient communication. *Proceedings of the National Academy of Sciences*, *108*(9), 3526–3529. (Cit. on p. 103).

Pickering, L., Corduas, M., Eisterhold, J., Seifried, B., Eggleston, A. & Attardo, S. (2009, November). Prosodic markers of saliency in humorous narratives. *Discourse Processes*, *46*(6), 517–540. (Cit. on p. 124).

Pipitone, R. N. & Gallup, G. G. (2008). Women's voice attractiveness varies across the menstrual cycle. *Evolution and Human Behavior*, *29*(4), 268–274. (Cit. on p. 102).

Pollet, T. V. & Nettle, D. (2008). Driving a hard bargain: Sex ratio and male marriage success in a historical us population. *Biology Letters*, *4*(1), 31–33. (Cit. on p. 113).

Polzehl, T. (2015). *Personality in speech: Assessment and automatic classification*. Springer. (Cit. on p. 99).

Poundstone, W. (1992). *Prisoner's Dilemma*. New York: Doubleday. (Cit. on pp. 46, 47).

Powers, A. & Kiesler, S. (2006). The advisor robot: Tracing people's mental model from a robot's physical attributes. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-Robot Interaction* (pp. 218–225). ACM. Salt Lake City, Utah, USA. (Cit. on p. 170).

Preston, D. R. (1999). *Handbook of perceptual dialectology*. John Benjamins Publishing. (Cit. on p. 33).

Preuschoft, S. (1992, April). “laughter” and “smile” in barbary macaques (macaca sylvanus). *Ethology*, *91*(3), 220–236. (Cit. on p. 123).

Przedlacka, J. (2008). Models and myth: Updating the (non)standard accents. In K. Dziubalska-Kołaczyk & J. Przedlacka (Eds.), *English pronunciation models: A changing scene* (2nd, pp. 17–35). Linguistic Insights: Studies in Language and Communication. Lang, Peter. (Cit. on p. 72).

Puts, D. A., Barndt, J. L., Welling, L. L. M., Dawood, K. & Burriss, R. P. (2011). Intrasexual competition among women: Vocal femininity affects perceptions of attractiveness and flirtatiousness. *Personality and Individual Differences*, *50*(1), 111–115. (Cit. on p. 113).

Puts, D. A., Gaulin, S. J. C. & Verdolini, K. (2006). Dominance and the evolution of sexual dimorphism in human voice pitch. *Evolution and Human Behavior*, *27*(4), 283–296. (Cit. on p. 102).

Puts, D. A., Hodges, C. R., Cárdenas, R. A. & Gaulin, S. J. C. (2007). Men’s voices as dominance signals: Vocal fundamental and formant frequencies influence dominance attributions among men. *Evolution and Human Behavior*, *28*(5), 340–344. (Cit. on pp. 102, 103, 114).

Quené, H. & Van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, *59*(4), 413–425. (Cit. on p. 112).

R Core Team. (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org. (Cit. on p. 57)

Ramirez, J. M. (2003, November). Hormones and aggression in childhood and adolescence. *Aggression and Violent Behavior*, *8*(6), 621–644. (Cit. on p. 102).

Rand, D. G. (2016). Cooperation, fast and slow meta-analytic evidence for a theory of social heuristics and self-interested deliberation. *Psychological Science*, 0956797616654455. (Cit. on pp. 47, 184, 186).

Rao, R. (2013, November). Prosodic consequences of sarcasm versus sincerity in mexican spanish. *Concentric: Studies in Linguistics*, *39*(2), 33–59. (Cit. on pp. 38, 76, 99).

Rau, P. L. P., Li, Y. & Li, D. (2009). Effects of communication style and culture on ability to accept recommendations from robots. *Computers in Human Behavior*, *25*(2), 587–595. (Cit. on pp. 64, 68, 80, 90, 128).

Re, D. E., O'Connor, J. J. M., Bennett, P. J. & Feinberg, D. R. (2012, March). Preferences for very low and very high voice pitch in humans. *PLoS ONE*, *7*(3), e32719. (Cit. on pp. 30, 31, 102, 106, 115).

Reed, L. I., Zeglen, K. N. & Schmidt, K. L. (2012, May). Facial expressions as honest signals of cooperative intent in a one-shot anonymous prisoner's dilemma game. *Evolution and Human Behavior*, *33*(3), 200–209. (Cit. on p. 139).

Reis, H. T., Wilson, I. M., Monestere, C., Bernstein, S., Clark, K., Seidl, E., . . . Radoane, K. (1990). What is smiling is beautiful and good. *European Journal of Social Psychology*, *20*(3), 259–267. (Cit. on pp. 122, 123, 139).

Repacholi, B., Slaughter, V., Pritchard, M. & Gibbs, V. (2003). Theory of mind, Machiavellianism, and social functioning in childhood. In B. Repacholi & V. Slaughter (Eds.), *Individual differences in theory of mind: Implications for typical and atypical development (macquarie monographs in cognitive science)* (pp. 67–97). New York, NY, US: Psychology Press. (Cit. on p. 26).

Reuter, M., Montag, C., Altmann, S., Bendlow, F., Elger, C., Kirsch, P., . . . Falk, A. (2009). Genetically determined differences in human trust behavior: The role of the oxytocin receptor gene. In *Neuropsychoeconomics conference proceedings* (Vol. 21). Bonn, Germany. (Cit. on pp. 25, 55).

Reysen, S. (2005). Construction of a new scale: The Reysen likability scale. *Social Behavior and Personality: an international journal*, *33*(2), 201–208. (Cit. on p. 156).

Rhodes, G. (2006). The evolutionary psychology of facial beauty. *Annual Review of Psychology*, *57*, 199–226. (Cit. on p. 30).

Riding, D., Lonsdale, D. & Brown, B. (2006). The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women. *Journal of Nonverbal Behavior*, *30*(2), 55–61. (Cit. on pp. 30, 102, 115).

Riedl, R. & Javor, A. (2012). The biology of trust: Integrating evidence from genetics, endocrinology, and functional brain imaging. *Journal of Neuroscience, Psychology, and Economics*, *5*(2), 63. (Cit. on pp. 15–17, 25, 26).

Roach, P. (1998). Some languages are spoken more quickly than others. In L. Bauer & P. Trudgill (Eds.), *Language myths* (pp. 150–158). Penguin. (Cit. on pp. 96, 117).

Rockwell, P., Buller, D. B. & Burgoon, J. K. (1997). The voice of deceit: Refining and expanding vocal cues to deception. *Communication Research Reports*, *14*(4), 451–459. (Cit. on p. 76).

Rosaldo, M. (1984). Toward an anthropology of self and feeling. *Culture theory: Essays on mind, self, and emotion*. (Cit. on p. 122).

Rosenberg, A. & Hirschberg, J. (2005). Acoustic/prosodic and lexical correlates of charismatic speech. In *Proceedings of INTERSPEECH* (pp. 513–516). Lisbon, Portugal. (Cit. on pp. 37, 76, 115, 116).

Rosenberg, A. & Hirschberg, J. (2009). Charisma perception from text and speech. *Speech Communication*, *51*(7), 640–655. (Cit. on p. 38).

Rothstein, B. & Uslaner, E. M. (2005). All for all: Equality, corruption, and social trust. *World politics*, *58*(01), 41–72. (Cit. on p. 23).

Rousseau, D. M., Sitkin, S. B., Burt, R. S. & Camerer, C. F. (1998). Not so different after all: A cross-discipline view of trust. *Academy of management review*, *23*(3), 393–404. (Cit. on p. 21).

Rule, N. O., Slepian, M. L. & Ambady, N. (2012). A memory advantage for untrustworthy faces. *Cognition*, *125*(2), 207–218. (Cit. on p. 25).

Russell, J. A., Bachorowski, J.-A. & Fernández-Dols, J.-M. (2003, February). Facial and vocal expressions of emotion. *Annual Review of Psychology*, *54*(1), 329–349. (Cit. on pp. 121, 122).

Sabatelli, R. M. & Rubin, M. (1986). Nonverbal expressiveness and physical attractiveness as mediators of interpersonal perceptions. *Journal of Nonverbal Behavior*, *10*(2), 120–133. (Cit. on p. 122).

Sailer, M., Hense, J. U., Mayr, S. K. & Mandl, H. (2017). How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction. *Computers in Human Behavior*, *69*, 371–380. (Cit. on p. 16).

Salem, M. & Dautenhahn, K. (2015). Evaluating trust and safety in HRI: Practical issues and ethical challenges. In *Emerging Policy and Ethics of Human-Robot Interaction: A Workshop at 10th ACM/IEEE International Conference on Human-Robot Interaction*. Portland, USA: ACM Press. (Cit. on p. 42).

Salem, M., Ziadee, M. & Sakr, M. (2013). Effects of politeness and interaction context on perception and experience of HRI. In *International Conference on Social Robotics* (pp. 531–541). Springer. Bristol, UK. (Cit. on pp. 15, 149, 166).

Sally, D. (1995). Conversation and cooperation in social dilemmas: A meta-analysis of experiments from 1958 to 1992. *Rationality and society*, *7*(1), 58–92. (Cit. on p. 183).

Samson, K. & Kostyszyn, P. (2015). Effects of cognitive load on trusting behavior – an experiment using the trust game. *PloS one*, *10*(5), e0127680. (Cit. on pp. 49, 187).

Sanfey, A. G. (2007). Social decision-making: Insights from game theory and neuroscience. *Science*, *318*(5850), 598–602. (Cit. on p. 46).

Sauter, D. A., Eisner, F., Ekman, P. & Scott, S. K. (2010). Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. *Proceedings of the National Academy of Sciences*, *107*(6), 2408–2412. (Cit. on p. 40).

Saxton, T. K., Caryl, P. G. & Roberts, C. S. (2006). Vocal and facial attractiveness judgments of children, adolescents and adults: The ontogeny of mate choice. *Ethology*, *112*(12), 1179–1185. (Cit. on pp. 30, 102, 115).

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J. & Frith, C. (2012, April). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, *7*(4), 413–422. (Cit. on p. 146).

Scassellati, B., Admoni, H. & Matarić, M. (2012). Robots for use in autism research. *Annual review of biomedical engineering*, *14*, 275–294. (Cit. on p. 170).

Scharlemann, J. P. W., Eckel, C. C., Kacelnik, A. & Wilson, R. K. (2001). The value of a smile: Game theory with a human face. *Journal of Economic Psychology*, *22*(5), 617–640. (Cit. on pp. 49, 119, 123, 139).

Scherer, K. R. (1987). Vocal assessment of affective disorders. In J. D. Maser (Ed.), *Depression and expressive behavior* (pp. 57–82). Lawrence Erlbaum Associates, Hillsdale, NJ. (Cit. on p. 121).

Scherer, K. R., Banse, R. & Wallbott, H. G. (2001, January). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-cultural psychology*, *32*(1), 76–92. (Cit. on p. 122).

Scherer, K. R., Johnstone, T. & Klasmeyer, G. (2003). Vocal expression of emotion. In R. J. Davidson, K. R. Scherer & H. H. Goldsmith (Eds.), *Handbook of affective sciences* (pp. 433–456). Series in Affective Science. Oxford University Press. (Cit. on p. 40).

Scherer, K. R., Ladd, D. R. & Silverman, K. E. A. (1984). Vocal cues to speaker affect: Testing two models. *The Journal of the Acoustical Society of America*, *76*(5), 1346–1356. Reference for reversed utterances in order to mask speech content. (Cit. on p. 118).

Scherer, K. R. & Wallbott, H. G. (1994). Evidence for universality and cultural variation of differential emotion response patterning. *Journal of Personality and Social Psychology*, *66*(2), 310–328. (Cit. on p. 122).

Schiel, F. (1999, August 7). Automatic phonetic transcription of non-prompted speech. In *Proceedings of the 14th International Congress of Phonetic Sciences* (pp. 607–610). S. (Cit. on p. 81).

Schimmack, U. & Diener, E. (2003). Predictive validity of explicit and implicit self-esteem for subjective well-being. *Journal of Research in Personality*, *37*(2), 100–106. (Cit. on p. 11).

Schug, J., Matsumoto, D., Horita, Y., Yamagishi, T. & Bonnet, K. (2010, March). Emotional expressivity as a signal of cooperation. *Evolution and Human Behavior*, *31*(2), 87–94. (Cit. on pp. 10, 122, 124, 140).

Schwarz, N. (1999). Self-reports: How the questions shape the answers. *American psychologist*, *54*(2), 93. (Cit. on pp. 11, 45).

Sell, A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C. & Gurven, M. (2009). Human adaptations for the visual assessment of strength and fighting ability from the body and face. *Proceedings of the Royal Society of London B: Biological Sciences*, *276*(1656), 575–584. (Cit. on p. 102).

Shamay-Tsoory, S. G. (2011). The neural bases for empathy. *The Neuroscientist*, *17*(1), 18–24. (Cit. on p. 28).

Shneiderman, B. & Plaisant, C. (2004). *Designing the User Interface: Strategies for Effective Human-Computer Interaction (4th Edition)*. Pearson Addison Wesley. (Cit. on p. 147).

Shor, R. E. (1978, January). The production and judgment of smile magnitude. *The Journal of General Psychology*, *98*(1), 79–96. (Cit. on p. 124).

Shue, Y.-L., Keating, P., Vicenik, C. & Yu, K. (2011, August 21). Voicesauce: A Program for Voice Analysis. In *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1846–1849). Hong Kong, China. (Cit. on p. 82).

Sims, V. K., Chin, M. G., Lum, H. C., Upham-Ellis, L., Ballion, T. & Lagattuta, N. C. (2009). Robots' auditory cues are subject to anthropomorphism. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, *18*, pp. 1418–1421). SAGE Publications. San Antonio, Texas, USA. (Cit. on p. 146).

Slonim, R. & Guillen, P. (2010). Gender selection discrimination: Evidence from a trust game. *Journal of Economic Behavior & Organization*, *76*(2), 385–405. (Cit. on pp. 39, 49, 77).

Smith, B. L., Brown, B. L., Strong, W. J. & Rencher, A. C. (1975). Effects of speech rate on personality perception. *Language and Speech*, *18*, 145–152. (Cit. on pp. 37, 76, 88, 99).

Smith, D. R. R., Patterson, R. D., Turner, R., Kawahara, H. & Irino, T. (2005). The processing and perception of size information in speech sounds. *The Journal of the Acoustical Society of America*, *117*(1), 305–318. (Cit. on p. 114).

Spence, K., Villar, G. & Arciuli, J. (2012). Markers of deception in Italian speech. *Frontiers in Psychology*, *3*, 453. (Cit. on p. 76).

Stanley, D. A., Sokol-Hessner, P., Fareri, D. S., Perino, M. T., Delgado, M. R., Banaji, M. R. & Phelps, E. A. (2012). Race and reputation: Perceived racial group trustworthiness influences the neural correlates of trust decisions. *Philosophical Transactions of the Royal Society B*, *367*, 744–753. (Cit. on p. 49).

Stanton, C. & Stevens, C. J. (2014). Robot pressure: The impact of robot eye gaze and lifelike bodily movements upon decision-making and trust. In *International Conference on Social Robotics* (pp. 330–339). Springer. Sidney, Australia. (Cit. on pp. 148, 168).

Staudte, M. & Crocker, M. W. (2011). Investigating joint attention mechanisms through spoken human–robot interaction. *Cognition*, *120*(2), 268–291. (Cit. on pp. 148, 165, 170).

Street, J., Richard L. (1984, December). Speech convergence and speech evaluation in fact-finding interviews. *Human Communication Research*, *11*(2), 139–169. (Cit. on pp. 37, 76, 96).

Street, J., Richard L. & Brady, R. M. (1982, December). Speech rate acceptance ranges as a function of evaluative domain, listener speech rate, and communication context. *Communication Monographs*, *49*(4), 290–308. (Cit. on pp. 37, 76).

Street, J., Richard L., Brady, R. M. & Putman, W. B. (1983, January). The influence of speech rate stereotypes and rate similarity or listeners' evaluations of speakers. *Journal of Language and Social Psychology*, *2*(1), 37–56. (Cit. on p. 96).

Streeter, L. A., Krauss, R. M., Geller, V., Olson, C. & Apple, W. (1977). Pitch changes during attempted deception. *Journal of Personality and Social Psychology*, *35*(5), 345–350. (Cit. on pp. 38, 75).

Strongman, K. T. & Woosley, J. (1967). Stereotyped reactions to regional accents. *British Journal of Social and Clinical Psychology*, *6*(3), 164–167. (Cit. on p. 34).

Surawski, M. K. & Ossoff, E. P. (2006). The effects of physical and vocal attractiveness on impression formation of politicians. *Current Psychology*, *25*(1), 15–27. (Cit. on p. 29).

Suzuki, A. & Suga, S. (2010, November). Enhanced memory for the wolf in sheep's clothing: Facial trustworthiness modulates face-trait associative memory. *Cognition*, *117*(2), 224–229. (Cit. on pp. 25, 179).

Syrdal, D. S., Dautenhahn, K., Woods, S. N., Walters, M. L. & Koay, K. L. (2007). Looking good? appearance preferences and robot personality inferences at zero acquaintance. In *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics* (pp. 86–92). (Cit. on p. 42).

Tager-Flusberg, H. & Sullivan, K. (2000). A componential view of theory of mind: Evidence from williams syndrome. *Cognition*, *76*(1), 59–90. (Cit. on p. 28).

Tamagawa, R., Watson, C. I., Kuo, I. H., MacDonald, B. A. & Broadbent, E. (2011). The effects of synthesized voice accents on user perceptions of robots. *International Journal of Social Robotics*, *3*(3), 253–262. (Cit. on p. 147).

Tartter, V. C. (1980, January). Happy talk: Perceptual and acoustic effects of smiling on speech. *Perception & Psychophysics*, *27*(1), 24–27. (Cit. on pp. 124, 127).

Tartter, V. C. & Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America*, *96*(4), 2101–2107. (Cit. on pp. 119, 125).

Thompson, D. F. & Meltzer, L. (1964). Communication of emotional intent by facial expression. *The Journal of Abnormal and Social Psychology*, *68*(2), 129–135. (Cit. on pp. 123, 140).

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C. & Feinberg, D. R. (2012). Voice pitch influences voting behaviour. *Evolution and Human Behavior*, *33*, 210–216. (Cit. on pp. 38, 77, 102, 113, 115).

Titze, I. R. (1989). Physiologic and acoustic differences between male and female voices. *The Journal of the Acoustical Society of America*, *85*(4), 1699–1707. (Cit. on p. 113).

Titze, I. R. (1994). *Principles of voice production*. Prentice Hall. (Cit. on p. 102).

Todorov, A. & Duchaine, B. (2008). Reading trustworthiness in faces without recognizing faces. *Cognitive Neuropsychology*, *25*(3), 395–410. (Cit. on p. 28).

Todorov, A., Pakrashi, M. & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. (Cit. on pp. 15, 26).

Todorov, A., Said, C. P., Engell, A. D. & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in cognitive sciences*, *12*(12), 455–460. (Cit. on p. 28).

Tomasello, M. (2008). *Origins of human communication*. MIT Press. (Cit. on p. 27).

Torre, I. (2014). Production and perception of smiling voice. In *Proceedings of the First Postgraduate and Academic Researchers in Linguistics at York conference (PARLAY 2013)*. York, UK. (Cit. on pp. ix, 119, 124, 125, 127).

Torre, I., Goslin, J. & White, L. (2015a). Investing in accents: How does experience mediate trust attributions to different voices? In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. Glasgow, U. K. (Cit. on pp. ix, 241).

Torre, I., Goslin, J. & White, L. (2015b, May 23). The effect of experience on trust attributions varies between two british accents. In *EPIP4: 4th International Conference on English Pronunciation: Issues & Practices* (pp. 146–148). (Cit. on p. ix).

Torre, I. & Loesche, F. (2016, December). Overcoming impasses in conversations: A creative business. *Creativity: Theories – Research – Applications*, *3*(2). (Cit. on pp. ix, 241).

Torre, I., White, L. & Goslin, J. (2016, May). Behavioural mediation of prosodic cues to implicit judgements of trustworthiness. In *Proceedings of the eighth International Conference on Speech Prosody 2016*. Boston, MA, USA: ISCA. (Cit. on pp. ix, 241).

Tortosa, M. I., Lupiáñez, J. & Ruz, M. (2013). Race, emotion and trust: An ERP study. *Brain Research*, *1494*, 44–55. (Cit. on p. 49).

Trivers, R. L. (1971). The evolution of reciprocal altruism. *The Quarterly review of biology*, *46*(1), 35–57. (Cit. on p. 23).

Trouvain, J., Schmidt, S., Schröder, M., Schmitz, M. & Barry, W. J. (2006, May 5). Modelling personality features by changing prosody in synthetic speech. In *Proceedings of the 3rd International Conference on Speech Prosody*. Dresden, Germany. (Cit. on pp. 39, 76, 85, 95, 99, 165).

Trudgill, P. (2000). *Sociolinguistics: An introduction to language and society*. Penguin UK. (Cit. on p. 31).

Tsantani, M. S., Belin, P., Paterson, H. M. & McAleer, P. (2016). Low vocal pitch preference drives first impressions irrespective of context in male voices but not in female voices. *Perception*, *45*(8), 946–963. (Cit. on pp. 113, 117).

Tversky, A. & Kahneman, D. (1971). Belief in the law of small numbers. In D. Kahneman, P. Slovic & A. Tversky (Eds.), *Judgment under uncertainty* (Vol. 76, *2*, pp. 23–31). Cambridge University Press. (Cit. on p. 186).

Tversky, A. & Kahneman, D. (1992, October). Advances in prospect theory: Cumulative representation of uncertainty. *Journal of Risk and Uncertainty*, *5*(4), 297–323. (Cit. on p. 96).

Uchanski, R. M., Choi, S. S., Braida, L. D., Reed, C. M. & Durlach, N. I. (1996). Speaking clearly for the hard of hearing iv: Further studies of the role of speaking rate. *Journal of Speech, Language, and Hearing Research*, *39*(3), 494–509. (Cit. on p. 104).

Uleman, J. S., Saribay, S. A. & Gonzalez, C. M. (2008). Spontaneous inferences, implicit impressions, and implicit theories. *Annual Review of Psychology*, *59*, 329–360. (Cit. on p. 16).

Valley, K. L., Moag, J. & Bazerman, M. H. (1998). A matter of trust':: Effects of communication on the efficiency and distribution of outcomes. *Journal of Economic Behavior & Organization*, *34*(2), 211–238. (Cit. on p. 183).

Van Borsel, G., De Cuypere, R., Rubens, B. & Destaerke, J. (2000). Voice problems in female-to-male transsexuals. *International Journal of Language & Communication Disorders*, *35*(3), 427–442. (Cit. on p. 102).

Van Kleef, G. A., De Dreu, C. K. W. & Manstead, A. S. R. (2010). An interpersonal approach to emotion in social decision making: The emotions as social information model. *Advances in Experimental Social Psychology*, *42*, 45–96. (Cit. on pp. 121, 124, 125, 139).

van Hooff, J. A. (1972). A comparative approach to the phylogeny of laughter and smiling. In R. A. Hinde (Ed.), *Non-verbal communication* (pp. 209–241). Cambridge University Press. (Cit. on p. 122).

Verhoeven, J., De Pauw, G. & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech*, *47*(3), 297–308. (Cit. on p. 103).

Villar, G., Arciuli, J. & Paterson, H. (2013). Vocal pitch production during lying: Beliefs about deception matter. *Psychiatry, Psychology and Law*, *20*(1), 123–132. (Cit. on pp. 38, 75).

Vrij, A., Edward, K., Roberts, K. P. & Bull, R. (2000). Detecting deceit via analysis of verbal and nonverbal behavior. *Journal of Nonverbal behavior*, *24*(4), 239–263. (Cit. on p. 76).

Vukovic, J., Feinberg, D. R., DeBruine, L., Smith, F. G. & Jones, B. C. (2010). Women's voice pitch is negatively correlated with health risk factors. *Journal of Evolutionary Psychology*, *8*(3), 217–225. (Cit. on p. 102).

Walters, M. L., Syrdal, D. S., Dautenhahn, K., Te Boekhorst, R. & Koay, K. L. (2008). 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. (Cit. on pp. 15, 147, 149, 166, 170).

Walters, M. L., Syrdal, D. S., Koay, K. L., Dautenhahn, K. & Te Boekhorst, R. (2008, August). Human approach distances to a mechanical-looking robot with different robot voice styles. In *RO-MAN 2008 - the 17th IEEE international symposium on robot and human interactive communication* (pp. 707–712). IEEE. Munich, Germany: IEEE. (Cit. on p. 170).

Wang, Z., Arndt, A. D., Singh, S. N., Biernat, M. & Liu, F. (2013). "you Lost Me at Hello": How and when accent-based biases are expressed and suppressed. *International Journal of Research in Marketing*, *30*, 185–196. (Cit. on p. 35).

Waytz, A., Morewedge, C. K., Epley, N., Monteleone, G., Gao, J.-H. & Cacioppo, J. T. (2010). Making sense by making sentient: Effectance motivation increases anthropomorphism. *Journal of personality and social psychology*, *99*(3), 410–435. (Cit. on p. 145).

Weber, J. M., Malhotra, D. & Murnighan, J. K. (2004). Normal acts of irrational trust: Motivated attributions and the trust development process. *Research in Organizational Behavior*, *26*, 75–101. (Cit. on p. 13).

Wells, J. C. (1982). *Accents of English*. Cambridge University Press. (Cit. on pp. 96, 117).

Wetzel, C. G., Wilson, T. D. & Kort, J. (1981). The halo effect revisited: Forewarned is not forearmed. *Journal of Experimental Social Psychology*, *17*(4), 427–439. (Cit. on pp. 28, 72).

Wicker, A. W. (1969). Attitudes versus actions: The relationship of verbal and overt behavioral responses to attitude objects. *Journal of Social Issues*, *25*(4), 41–78. (Cit. on pp. 11, 45).

Wickham, H. (2009). *Ggplot2: Elegant graphics for data analysis*. Springer-Verlag New York. (Cit. on p. 57).

Wilkinson, A. M. (1965). *Spoken english*. Educational review. 17/2: Occasional publications. University of Birmingham. (Cit. on pp. 34, 72, 92, 98).

Wilkinson, G. S., Carter, G. G., Bohn, K. M. & Adams, D. M. (2016). Non-kin cooperation in bats. *Phil. Trans. R. Soc. B*, *371*(1687), 20150095. (Cit. on p. 23).

Willinger, M., Keser, C., Lohmann, C. & Usunier, J.-C. (2003). A comparison of trust and reciprocity between france and germany: Experimental investigation based on the investment game. *Journal of Economic Psychology*, *24*(4), 447–466. (Cit. on p. 49).

Willis, J. & Todorov, A. (2006). First impressions: Making up your mind after a 100-ms exposure to a face. *Psychological Science*, *17*(7), 592–598. (Cit. on pp. 14, 15, 24, 37, 59, 104).

Wolff, S. E. & Puts, D. A. (2010). Vocal masculinity is a robust dominance signal in men. *Behavioral Ecology and Sociobiology*, *64*(10), 1673–1683. (Cit. on pp. 102, 103, 114).

Wood, R. M., Rilling, J. K., Sanfey, A. G., Bhagwagar, Z. & Rogers, R. D. (2006). Effects of tryptophan depletion on the performance of an iterated prisoner's dilemma game in healthy adults. *Neuropsychopharmacology*, *31*(5), 1075–1084. (Cit. on p. 25).

Wout, M. V. & Sanfey, A. G. (2008). Friend or foe: The effect of implicit trustworthiness judgments in social decision-making. *Cognition*, *108*(3), 796–803. (Cit. on p. 48).

Wright, S. C., Aron, A., McLaughlin-Volpe, T. & Ropp, S. A. (1997). The extended contact effect: Knowledge of cross-group friendships and prejudice. *Journal of Personality and Social Psychology*, *73*(1), 73. (Cit. on pp. 14, 104).

Xu, Y., Lee, A., Wu, W.-L., Liu, X. & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PloS one*, *8*(4), e62397. (Cit. on pp. 30, 40, 77, 102, 114).

Yamagishi, T., Tanida, S., Mashima, R., Shimoma, E. & Kanazawa, S. (2003). You can judge a book by its cover: Evidence that cheaters may look different from cooperators. *Evolution and Human Behavior*, *24*(4), 290–301. (Cit. on p. 24).

Yamamoto, Y. (1990). A morality based on trust: Some reflections on Japanese morality. *Philosophy East and West*, *40*(4), 451–469. (Cit. on p. 22).

Zak, P. J. (2003). Trust. *Capco Institute Journal of Financial Transformation*, *7*, 17–24. (Cit. on p. 54).

Zak, P. J. & Fakhar, A. (2006, December). Neuroactive hormones and interpersonal trust: International evidence. *Economics & Human Biology*, *4*(3), 412–429. (Cit. on p. 23).

Zak, P. J. & Knack, S. (2001). Trust and growth. *The economic journal*, *111*(470), 295–321. (Cit. on p. 23).

Zak, P. J., Kurzban, R. & Matzner, W. T. (2005, December). Oxytocin is associated with human trustworthiness. *Hormones and Behavior*, *48*(5), 522–527. (Cit. on p. 25).

Zanatto, D., Patacchiola, M., Goslin, J. & Cangelosi, A. (2016, March). Priming anthropomorphism: Can the credibility of humanlike robots be transferred to non-humanlike robots? In *2016 11th ACM/IEEE international conference on human-robot interaction (HRI)* (pp. 543–544). IEEE Press. Christchurch, New Zealand: IEEE. (Cit. on p. 148).

Zebrowitz, L. A. & Collins, M. A. (1997, August). Accurate social perception at zero acquaintance: The affordances of a gibsonian approach. *Personality and Social Psychology Review*, *1*(3), 204–223. (Cit. on pp. 14, 104).

Zebrowitz, L. A. & Montepare, J. M. (2008, May). Social psychological face perception: Why appearance matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. (Cit. on p. 28).

Zhang, T., Zhu, B., Lee, L. & Kaber, D. (2008, August). Service robot anthropomorphism and interface design for emotion in human-robot interaction. In *2008 IEEE international conference on automation science and engineering* (pp. 674–679). IEEE. Marriott Key Bridge Arlington, VA, USA: IEEE. (Cit. on p. 170).

Zipf, G. K. (2012, June 6). *Human behavior and the principle of least effort*. First published in 1949. Martino Fine Books. (Cit. on p. 103).

Złotowski, J., Strasser, E. & Bartneck, C. (2014). Dimensions of anthropomorphism: From humanness to humanlikeness. In *Proceedings of the 2014 ACM/IEEE international conference on human-robot interaction - HRI '14* (pp. 66–73). ACM. Bielefeld, Germany: ACM Press. (Cit. on p. 145).

Zuckerman, M., DeFrank, R. S., Hall, J. A., Larrance, D. T. & Rosenthal, R. (1979, July). Facial and vocal cues of deception and honesty. *Journal of Experimental Social Psychology*, *15*(4), 378–396. (Cit. on pp. 38, 75, 99).

Zuckerman, M. & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, *13*(2), 67–82. (Cit. on pp. 28, 30).

Zuckerman, M. & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior*, *17*(2), 119–135. (Cit. on pp. 29, 101).

Bound copies of published papers

The next pages show copies of the following published materials:

Torre, I., Goslin, J. & White, L. (2015a). Investing in accents: How does experience mediate trust attributions to different voices? In *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS 2015)*. Glasgow, U. K..

Torre, I. & Loesche, F. (2016, December). Overcoming impasses in conversations: A creative business. *Creativity: Theories – Research – Applications*, *3*(2).

Torre, I., White, L. & Goslin, J. (2016, May). Behavioural mediation of prosodic cues to implicit judgements of trustworthiness. In *Proceedings of the eighth International Conference on Speech Prosody 2016*. Boston, MA, USA: ISCA.

Investing in accents: How does experience mediate trust attributions to different voices?

Ilaria Torre, Jeremy Goslin, Laurence White

School of Psychology, Plymouth University
ilaria.torre@plymouth.ac.uk; jeremy.goslin@plymouth.ac.uk; laurence.white@plymouth.ac.uk

ABSTRACT

Speakers' accents have been claimed to influence initial judgements of personality traits, such as trustworthiness. We examined how personal experience with specific accents may serve to modify initial trust attributions, using an iterated trust game in which participants make investments with virtual players. The virtual player's accent was either Liverpool English or Standard Southern British English (SSBE), and they systematically returned investments either generously or meanly. When the virtual player was generous, participants consistently invested more with the SSBE-accented player throughout the game. When the virtual player was mean, participants initially invested more with SSBE, but after a few rounds the pattern reversed, and they subsequently invested more with the Liverpool-accented player, even though the pattern of investment returns between accents was the same. This interaction suggests that initial voice-based personality attributions may mediate the interpretation of a speaker's subsequent behaviour.

Keywords: English accents, game theory.

1. INTRODUCTION

We make initial judgements about personality based on very limited evidence, such as seeing a face for 100ms, or hearing the word "hello" [20, 13]. Many physical and behavioural characteristics (e.g., facial appearance and expression, dress, voice quality, accent) contribute to the formation of these impressions. It is still unclear, however, what specific vocal characteristics contribute to attributions of personality traits. Indeed – considering specifically trustworthiness – existing research is somewhat contradictory about the acoustic characteristics that give rise to trust judgments. For example, in a mock election scenario, male and female participants typically voted for candidates with lower-pitched voice [18]. Supporting this, higher pitch early in interactions predicts lower trust levels [8]. This effect diminishes over time, however [8], and smiling (which tends to shorten the vocal tract and raise pitch [17]) can be perceived in the voice and is likely to increase trust [8]. Furthermore, although gender strongly impacts on overall pitch level, neither gender consistently evokes higher trust judgments (e.g. [3, 5, 14, 16]).

Accent differences can also suggest personality stereotypes even to non-native speakers [11, 1], and native accents may be perceived as more trustworthy than non-native accents [12]. Accents, though, are often intrinsically related to geographic regions (with the exception of accents associated with social class, such as Standard Southern British English – SSBE), and stereotypes based on general socio-economic perceptions of particular regions may impact on personality attributions. Some studies have shown that standard accents such as SSBE are rated as more pleasant and attractive than, for example, city accents such as Liverpool or Birmingham [2, 7, 9]. People may not actually be very effective at localising accents however [1], and it is possible that some vocal characteristics of accents may mediate trust judgements independent of regional stereotypes. Furthermore, most research on accent attributions focuses on immediate impressions, without taking into account how the attribution might evolve over time according to the speaker's behaviour. We tried to explore the dynamic impact of voice on trust attributions with a novel experimental design based on game theory. We used speakers of two British accents – SSBE and Liverpool – which have been suggested to evoke contrasting trust attributions [2, 9], in an experiment where participants have to make monetary investments with virtual players.

2. METHOD

We used an iterated trust game – "the investment game" – to test how voice-based trust attributions change with experience.

2.1. The investment game

Based in game theory [15], some studies have employed a game scenario requiring simulated monetary investments to investigate trust attributions to a range of characteristics, including gender [3, 4, 6], race and emotion [19], and facial expressions [10]. To our knowledge, however, such games have never been applied to study trust attributions to different voices. We used an iterated investment game to test the hypothesis that the initial trust attribution to two accents (SSBE and Liverpool English) changes over time, according to the way the accented virtual player behaves.

2.1.1. Participants

There were 44 native English participants (35 females, 9 males) aged 18-45 (mean 22.05). They were university undergraduate students who received course credit for participation. Self-reports on participants' geographical origins were: southwest England (n = 31), southeast England (n = 9), Midlands (n = 2), Scotland (n = 1) and Wales (n = 1).

2.1.2. Stimuli

The utterances used in the game and the subsequent questionnaire were recorded in a sound-attenuated booth. We recorded seven male native British English speakers, from Liverpool, Edinburgh, Birmingham, south London, Huddersfield and Bournemouth, plus an SSBE speaker, and five male second language (L2) English speakers, from Germany (Saxony), France (Normandy), Italy (Tuscany), Greece (Macedonia) and India (National Capital Region). Each speaker read two blocks of 20 sentences (one for each round of the game, described below), all approximately the same length (mean number of syllables per sentence 16.6, SD 1.08). Apart from the first utterance of each block, which served for the virtual player to introduce himself, all other utterances were about strategies to follow in the game, e.g.: "I'm going to return more money now, if you invest more as well"; "Remember, there is potential for earning, if we both trust each other"; "The goal of the game is to earn as much money as possible". The SSBE and Liverpool utterances were used in the game, while the others were used in the questionnaire completed at the end. The recorded utterances were amplitude-normalized, and a noise-removal filter was applied.

2.1.3. Procedure

Participants were told that the goal of the game was to earn as much money as possible, and that mutual co-operation with the other (virtual) player would lead to greater profit. They were informed that they could not verbally interact with the other player, but that they would hear an utterance spoken by him at the beginning of each round. The first player (the participant) started with a notional sum of £8 at the beginning of each of the 20 rounds. He/she then decided whether to invest all, part, or none of it with a virtual player, who then received three times the invested amount. The latter was programmed to have one of two behaviours, either returning 120% to 240% of the invested money to the participant (*generous* condition) or 0% to 120% (*mean* condition). Thus, in the *generous* condition, if the participants invested some of the money they were given, they would end the round with more money than they started with. The virtual player's utterances within each game were always in the same (SSBE or Liverpool) accent. Altogether, there were four accent-return pairings (Liverpool-*generous*, Liverpool-*mean*, SSBE-*generous*, SSBE-*mean*). Each participant engaged in two games, one for each accent, and one for each behaviour (*generous*/*mean*), with a different set of 20 sentences heard for each version of the game. Each round of the game proceeded as follows: participants heard the utterance from the virtual player; they indicated how much of their £8 they wished to invest (by pressing a digit key); they saw a summary screen with all the "monetary" transactions to and from the virtual player that had happened during the round, including the return on their investment.

2.2. Questionnaire

2.2.1. Procedure

After the participants finished the two games, they completed a questionnaire, during which they heard two utterances from each of the 12 speakers (see 2.1.2) in random order, and were asked to rate the speaker's voice on a 7-point Likert scale (1 = very untrustworthy, 7 = very trustworthy).

3. RESULTS AND DISCUSSION

3.1. Investment game

The overall investment pattern was dictated by the virtual player's behaviour, with participants investing consistently more with the *generous* virtual player (Fig. 1), regardless of his accent (Fig. 2). A paired t-test confirmed that investments in the *generous* condition were higher than in the *mean* condition ($t(879) = 32.82$, $p < 0.001$). There was no difference in the overall investments made according to the accent of the virtual player ($t(879) = -0.42$, $p = 0.68$).

Figure 1: Mean investments in the *generous* and *mean* conditions.

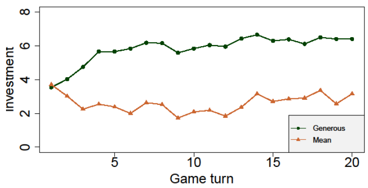
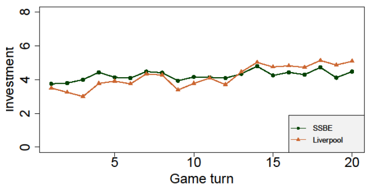


Figure 2: Mean investments in the SSBE and Liverpool accent conditions.



A linear mixed-effects model was fitted to the data, with investment as dependent variable, accent, return behaviour and game turn order as independent variables, and participants as a random factor. The model showed a main effect of the return condition ($\chi^2(1) = 983.36$, $p < 0.001$), no effect for accent ($\chi^2(1) = 0.48$, $p = 0.49$), a significant interaction between accent and return condition ($\chi^2(1) = 7.38$, $p = 0.007$), a main effect of game turn ($\chi^2(1) = 65.45$, $p < 0.001$) and a significant three-way interaction ($\chi^2(1) = 12.39$, $p < 0.001$).

Given the interactions, we then considered the two return behaviour conditions separately. In the *generous* condition (Fig. 3), there was a main effect of accent ($\chi^2(1) = 4.93$, $p = 0.026$), with higher investment to SSBE, a main effect of game turn ($\chi^2(1) = 151.77$, $p < 0.001$), with an overall increase in investment as the game proceeded, but no significant interaction ($\chi^2(1) = 1.11$, $p = 0.29$). For the *mean* condition (Fig. 4), the model showed a main effect of accent ($\chi^2(1) = 4.94$, $p = 0.026$), with higher overall investment for Liverpool, no effect of game turn ($\chi^2(1) = 1.81$, $p = 0.18$), and a significant interaction of accent and game turn ($\chi^2(1) = 36.44$, $p < 0.001$).

Thus, differences between the responses to the accents emerge when considering the two behaviour conditions separately. When the virtual player was generous, participants consistently invested more with the SSBE-accented player throughout the game (Fig. 3), supporting findings of relative trustworthiness of SSBE [2, 9]. When the virtual player was mean, participants initially invested more with SSBE, but after three rounds the pattern reversed, and they subsequently invested more with the Liverpool-accented player (Fig. 4), even though the pattern of investment return between accents was the same.

Figure 3: Mean investments to SSBE and Liverpool accents in the *generous* condition.

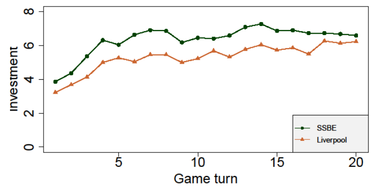
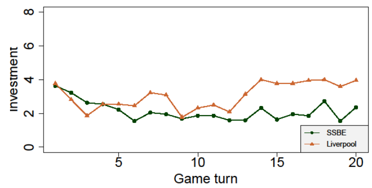


Figure 4: Mean investments to SSBE and Liverpool accents in the *mean* condition.

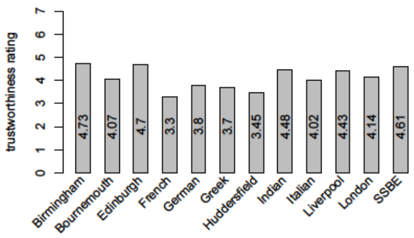


Using an implicit test of trustworthiness, these preliminary results (based on a single-speaker per accent) support preconceptions that accent-personality stereotypes exist. When the virtual player was behaving generously, participants apparently trusted the SSBE accent more. However, when the virtual player consistently returned little of the invested money, participants invested less in the SSBE-accented player. This suggests that speakers judged initially more reliable may be more severely discredited if their behaviour is at odds with first impressions.

3.2. Trustworthiness questionnaire

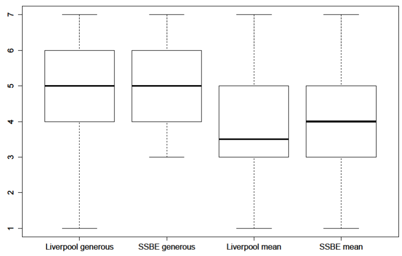
Fig. 5 shows the mean trustworthiness ratings of the 12 sampled speakers. The pattern of ratings differed somewhat from previous trust attributions to accents of English (e.g. [2, 9]), where, for example, Birmingham accents scored relatively low and Yorkshire accents scored much higher. In this preliminary study, however, there was only a single speaker representing each region, and idiolectal characteristics clearly affected ratings in addition to any regional associations.

Figure 5: Mean trustworthiness ratings of 12 English speakers.



The ratings for the SSBE and Liverpool voices are likely to be influenced by participants' foregoing experiences in the investment game. Fig. 6 shows the trustworthiness ratings of SSBE and Liverpool, grouped by the two return conditions. Ratings in the Liverpool-*generous* were higher than in the Liverpool-*mean* condition ($t(42) = 2.51$, $p < 0.05$), whilst the difference between the SSBE-*generous* and SSBE-*mean* condition approached significance ($t(40) = 1.87$, $p = 0.07$). Thus, it appears that overall the explicit trust ratings are in line with participants' investment behaviour.

Figure 6: Boxplot showing the trustworthiness ratings of SSBE and Liverpool in the two behaviour conditions.



It is worth noting that, of the 44 participants in the experiment, only four correctly identified the provenance of the Liverpool accent. Thus, social stereotypes may play a minor role in trust attributions in this study compared to idiolectal characteristics.

4. CONCLUSION

In the *generous* condition, there was higher investment with the SSBE speaker, somewhat reinforcing previous findings regarding accent-trust attributions. More interestingly, the accent of the speaker strongly affected how participants reacted to negative behaviour on the part of the virtual player. The speaker with the non-regional (standard) accent initially attracted higher investment, but showed a greater drop in investment than the regional speaker once the negative pattern of returns became evident. This preliminary result intriguingly indicates that socially "prestigious" accents may incur a more negative response to perceived unjust behaviour.

Data on accent attributions gathered with traditional sociolinguistic methods, such as questionnaires, only provide subjective, static measures. The novel methodology used in this study provides an implicit measure of trust; furthermore, it allows us to study how trust attributions change over time. Thus, it is particularly useful for studying the effect of observed behaviour on implicit personality attributions. However, the game data so far uses only two voices, and further studies will need to expand the range of voices to test how general is the observed contrast in investment between accents in the negative behaviour scenario. For example, it would be interesting to see whether other relatively "prestigious" accents incur similar investment penalties once initial stereotypes are undermined by behaviour. We also need to determine the degree to which trust attributions are due to individual differences in voice quality and prosody. The present study is part of a broader research programme investigating voice characteristics that influence trust attributions.

5. AKNOWLEDGEMENTS

This work was supported by CogNovo (FP7-PEOPLE-2013-ITN-604764), a project funded by the EU Marie Curie programme.

6. REFERENCES

[1] Bayard, D., Gallois, C., Weatherall, A., Pittam, J. (2001). Pax Americana? Accent attitudinal evaluations in New Zealand, Australia and America. *Journal of Sociolinguistics 5(1)*, 22-49.

[2] Bishop, H., Coupland, N., Garrett, P. (2005). Conceptual accent evaluation: Thirty years of accent prejudice in the UK. *Acta Linguistica Hafniensia: International Journal of Linguistics, 37(1)*, 131-154.

[3] Boenin, A., Serra, D. (2009). Gender pairing bias in trustworthiness. *The Journal of Socio-Economics 38*, 779–789.

[4] Buchan, N. R., Croson, R. T. A., Solnick, S. (2008). Trust and gender: An examination of behavior and beliefs in the Investment Game. *Journal of Economic Behavior & Organization 68*, 466–476.

[5] Chaudhuri, A., Gangadharan, L. (2007). An experimental analysis of trust and trustworthiness. *Southern Economic Journal 73(4)*, 959-985.

[6] Chaudhuri, A., Paichayontvijit, T., Shen, L. (2013). Gender differences in trust and trustworthiness: Individuals, single sex and mixed sex groups. *Journal of Economic Psychology 34*, 181–194.
[7] Dixon, J. A., Mahoney, B., Cocks, R. (2002). Accents of guilt? Effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology 21 (2)*, 162-168.
[8] Elkins, A. C., Derrick, D. C. (2013). The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents. *Group Decision and Negotiation 22 (5)*, 897-913.
[9] Fuertes, J. N., Gottdiener, W. H., Martin, H., Gilbert, T. C., Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology 42*, 120–133.
[10] Krumhuber, E., Manstead, A. S. R., Cosker, D., Marshall, D., Rosin, P. L., Kappas, A. (2007). Facial dynamics as indicators of trustworthiness and cooperative behavior. *Emotion 7 (4)*, 730–735.
[11] Ladegaard, H. J. (2008). National stereotypes and language attitudes: the perception of British, American and Australian language and culture in Denmark. *Language & Communication 18*, 251-274.
[12] Lev-Ari, S., Keysar, B. (2008). Why don't we believe non-native speakers? The influence of accent on credibility. *Journal of Experimental Social Psychology 46*, 1093–1096.
[13] McAleer, P., Todorov, A., Belin, P. (2014). How do you say 'Hello'? Personality impressions from brief novel voices. *PLoS ONE 9(3)*: e90779.
[14] Nass, C., Brave, S. (2005). *Wired for speech: How voice activates and advances the human-computer relationship*. Cambridge, MA: MIT Press.
[15] Poundstone, W. (1992). *Prisoner's dilemma*. New York: Doubleday.
[16] Slonim, R., Guillen, P. (2010). Gender selection discrimination: Evidence from a Trust game. *Journal of Economic Behavior & Organization 76*, 385–405.
[17] Tartter, V. C. and Braun, D. (1994). Hearing smiles and frowns in normal and whisper registers. *Journal of the Acoustical Society of America 96 (4)*, 2101-2107.
[18] Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., Feinberg, D. R. (2012). Voice pitch influences voting behaviour. *Evolution and Human Behavior 33*, 210–216.
[19] Tortosa, M. I., Lupiáñez, J., Ruz, M. (2013). Race, emotion and trust: An ERP study. *Brain Research 1494*, 44–55.
[20] Willis, J., Todorov, A. (2006). First impressions: making up your mind after a 100-ms exposure to a face. *Psychological Science 17(7)*, 592-8.

DE GRUYTER OPEN

CREATIVITY

Vol. 3, Issue 2, 2016

Theories – Research – Applications

Overcoming Impasses in Conversations: A Creative Business*

Ilaria Torre
Plymouth University, UK
E-mail address: Ilaria.Torre@Plymouth.ac.uk

Frank Loesche
Plymouth University, UK
E-mail address: Frank.Loesche@Plymouth.ac.uk

A R T I C L E I N F O

Keywords:
Impasse
Conversation
Creative problem solving
Conversation Analysis

Article history:
Received: 13 March 2016
Received in revised form: 29 July 2016
Accepted: 15 October 2016
ISSN: 2354-0036
DOI: 10.1515/ctra-2016-0016

A B S T R A C T

The ability to communicate complex meanings is a specific human ability which plays a crucial role in social interactions. A habitual example of these interactions is conversation. However, we observe that spontaneous conversation often hits an *impasse* when none of the interlocutors immediately produces a follow-up utterance. The existence of impasses in conversations, and the way that interlocutors overcome them provide evidence for our argument that conversation is a sequence of creative problem solving. In this work we use techniques from Conversation Analysis (CA) on publicly available databases of naturally-occurring speech and we suggest a framework to understand how impasses are reached and overcome. As a result, we hope to reveal yet another instance of the bond between language and creativity.

INTRODUCTION

In this paper we discuss conversations between humans and explore the possibility of interpreting them as continuous creative problem solving, where the problem is represented by the common phenomenon of impasse.

Conversation

While scholars have not reached an agreement on a formal definition of conversation, it is at least possible to isolate the phenomenon according to the elements associated with it. A conversation is often referred to as a meaningful verbal communication between at least two participants (Donaldson, 1979), who cooperatively construct an interaction (Clark, 1996; Warren, 2006). Ritualized exchanges as well as interactions between obvious status differences are often excluded from the definitions (Donaldson, 1979; War-

* This work was supported by CogNovo (FP7-PEOPLE2013-ITN-604764), a project funded by the EU Marie Curie programme. The authors wish to thank Dr. Traci Walker for her suggestions on the conversation corpus

ren, 2006). In this sense, interactions such as greetings, lectures, or orders are not usually considered conversations. Conversation generally serves the purpose of creating and maintaining social relationships (Donaldson, 1979; Kaplan, 1997; Sacks, Schegloff, & Jefferson, 1974). Its social and collaborative nature can also be evinced by one of its most prominent and necessary features, turn-taking. If interlocutors were not able to follow the implicit *rules* of turn-taking, it would not be a conversation, but rather a chaotic ensemble of sounds. Even though the content of a conversation might not be an information exchange, the fact that each utterance is an action which implies something is a well-known theory in Philosophy of Language (Austin, 1975; Grice, 1957, 1975; Searle, 1969). From this point of view, we can say that conversations are meaningful because, even though they might not convey any new information to the interlocutors, they mean something for all of them, and they can be the means by which a change in the environment takes place (e.g. someone says "I'm cold" followed by another person closing the window). Another element associated with conversation and that distinguishes it, for example, from written communication is improvisation: interlocutors in a conversation do not know in advance what others or themselves will say in the next turns, so they need to be able to - as Donaldson puts it - creatively adapt to it (Donaldson, 1979; Halliday, 1994; McCarthy, 2001). Thus, following the summary in Jordan et al. (2009), we can conclude that conversation is made up of three main components: social collaboration, meaningfulness and improvisation.

Problem Solving

Problem solving can be seen as a way of reaching a goal. A problem can be well-defined, i.e. having an expected solution, a clear goal, and a defined path on how to reach it. Lacking any of these elements makes it ill-defined. However, the boundary between these types of problem does not seem to be easily distinguishable (Reed, 2015; Simon, 1973). Creative problem solving adds a notion of creativity to it: Runco & Jaeger emphasize that a creative solution has to meet the minimal requirement of being novel work that is accepted as tenable or useful or satisfying by a group in time, as initially put forward by Stein (Runco & Jaeger, 2012; Stein, 1953). Often a third attribute of surprise is considered to be compulsory (Boden, 2004). These three attributes have different names in the creativity and creative problem solving literature, either using synonyms or similar but not equal terms. In social situations additional elements are frequently mentioned as identifiers of creativity - such as expression - while usefulness might be referred to as social appreciation as a result of social evaluation (Fischer, Giaccardi, Eden, Sugimoto, & Ye, 2005). Taking the personal-psychological creativity of the individual into account (see Boden, 2004), a distinction between convergent and divergent thinking, following

Guilford, is helpful: convergent tasks have one possible solution while divergent thinking might arrive at several correct solutions (Guilford, 1967). Both processes have been linked to problem solving, for example as a search within a problem space (e.g. Newell & Simon, 1972), or as a divergent production of potential solutions (Gilhooly, Fioratou, Anthony, & Wynn, 2007).

Conversation as Creative Problem Solving

In the rest of the section we explore similarities between conversation and creative problem solving. As we have seen, conversation includes elements of social collaboration, meaningfulness, and improvisation, while creative problem solving in a social situation can be described as a novel, expressive, and useful way of reaching a goal. In problem solving this goal does not have to be tangible, e.g. for ill-defined problems. Just as in Fischer et al.'s (2005) description of social creativity people need to express their ideas, interlocutors articulate their utterances in a conversation - they express them. Similarly, the other attributes mentioned above show a large potential overlap between conversation and creative problem solving: meaningfulness, as described by Jordan et al. (2009), shows some similarity to the idea of usefulness. Improvisation, either as an acquired skill or a constrained combinatorial creation, includes subjective novelty in this context. Since conversations usually do not aim at a clearly defined goal, and since their structure is also not defined, they show similarities to ill-defined problems. Similarly, Demuth and Glaveanu (2016) argue that spontaneity and creativity contribute to the construction of utterances based on predictions of what the other person might say in the future; in this sense, conversation can also be understood as a *problem finding* approach in an ill-defined environment.

The link between creativity and conversation can be observed on a more granular level as well. For example, Kaufman and Beghetto (2013) distinguish between subjective self-discovery (mini-C), everyday creativity (little-c), and geniuses (big-C). Between and within each of these groups different tasks might require various time spans to be solved. Similarly, a longer conversation can consist of smaller parts, each of which can be dedicated to a certain topic or carry a distinguishable meaning. Each of these sub-parts shows the same previously highlighted characteristics of the whole conversation: it is a social collaboration, which includes novelty and improvisation, and it is useful by being meaningful for the participants. As these sub-parts sequentially follow one another, eventually either another sub-part is initiated or the whole conversation is terminated after the end of a sub-part. We should mention that we are not considering the ability to create new linguistic symbols, which is another instance of creativity in language, but we only focus on the process of overcoming impasses.

Impasse

Creative problem solving, or more generally creativity, is often described using a sequential model similar to the famous one proposed in Wallas (1926). He suggested that creative problem solving can be understood as a sequence of phases which he named preparation, incubation, illumination, and verification. A following phase can only start if the previous one is completed. Even though the number and nature of these sequential steps has been the subject of debate for the past 90 years, the idea of using sequential or componential models has proven to be a valid way of describing the creative process in individual as well as social contexts (also see Amabile, 1983). Howard, Culley, & Dekoninck (2007) compare 42 models, more than 20 of which present a sequence of phases similar to those previously mentioned. A situation within this sequence, *where one does not know how to proceed*, is referred to as an impasse (Cranford & Moss, 2012). An impasse is therefore, at least from an external point of view, accompanied by the passing of time, without any advancements in the process, or, to use terms that are found in conversation studies, a gap or lapse (Sacks, Schegloff, & Jefferson, 1974). Consequently, all the sequential models mentioned by Howard, Culley, and Dekoninck (2007) include a concept similar to that of an impasse.

Framework

In this section we want to offer a more detailed speculation on what might happen during a conversation regarding impasses, but reduce our explanation to the elements of a conversation necessary to follow this article. Since elements of our model relate to Conversation Analysis (CA) as well as creative problem solving, we avoid technical terms in order not to confuse them. In spoken communication silences are intentionally used as a rhetorical instrument and to prevent or even provoke turn-taking, but they also occur unplanned by the interlocutors. In this paper we focus on the latter case and refer to these occurrences as impasses.

To start with some examples: during a conversation, a difficult and open-ended question is posed, but no-one has an answer to it, or a topic in a conversation has been exhausted. In both cases, the interlocutors might have difficulties finding the next meaningful and non-repetitive thing to say. If one of them starts a follow-up topic early enough, the others might still have experienced the impasse, even though it might have gone unnoticed by an external listener. On the other hand, if none of the speakers is able to proceed, this impasse produces an unnaturally long silence - what Sacks, Schegloff, and Jefferson (1974) call a *lapse*. We argue that these pauses can be observed by an outside listener, especially since Stivers et al. (2009) have shown universal patterns in response latency, i.e. silences, across languages and cultures. From a probabilistic point of view, the observable impasses are more likely to happen with a low number of interlocutors, with the additional benefit of having a simpler overall communication structure.

To illustrate the context of an impasse happening within a 2-interlocutor conversation, we refer to Figure 1.

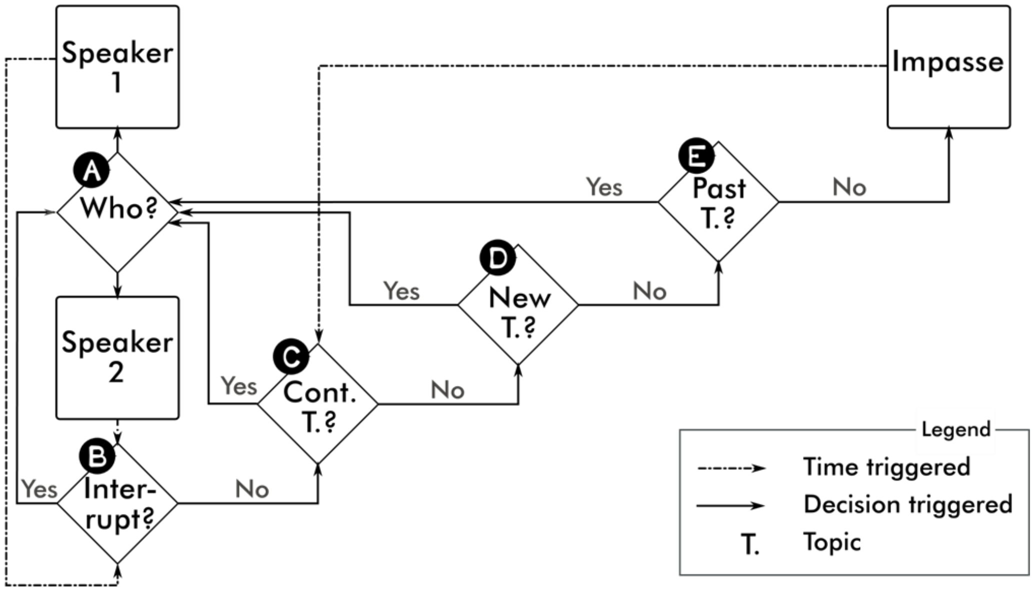


Figure 1 Impasses in conversations

All the square boxes represent time-based events that can be observed from an outside perspective. One of them represents an impasse, as we hypothesize that solving it takes considerable time that can be observed as an increased response latency by an outside observer. The passing time is emphasized in our figure by the time triggered transitions to one of the decisions, represented by diamond shapes (e.g. B). These decisions, on the other hand, happen quickly between interlocutors (e.g. A and B), or internally to one speaker (C to E) - too fast to be identified by outside observers. Notice that, as ritualized communication is excluded from our analysis, following the definitions described in the Introduction, elements such as greetings and closing remarks are not included in the diagram. Step A represents the moment when either Speaker 1 or Speaker 2 is starting to talk. For an extensive discussion on this step and the processes involved, we refer to Sacks, Schegloff, and Jefferson (1974), Selting (1996), and Stivers et al. (2009). For our model it is sufficient to have a step A which results in one speaker starting to talk. Once this decision is made, the active speaker talks for a certain amount of time, checking for cues of being interrupted by the other speaker at B. This can happen very quickly, as shown by Chambliss and Feeny (1992), and Cummins, Li, and Wang (2013). If the speaker understands that he / she is being interrupted, the turn is either taken by the other speaker or the same speaker continues at A. In the latter case and if the topic is not yet depleted and can be continued (C), the conversation is continued at A. If the topic cannot be continued at C and one of the speakers has a new topic ready, the conversation continues from D to A. The conversation also continues in the case of one of the speakers deciding to revisit a previous topic at E. If this is not the case either, the conversation does not progress and arrives at an impasse. During the impasse a speaker might re-evaluate the possibility of either continuing the current topic at C, starting a new one at D, or going back to a previous topic at E. Participants might iterate through these possibilities for some time before one of them is able to start talking again at A. If the step from C to D is iterative and stays within the semantically constrained problem space, these loops are similar to the search described in Newell and Simon (1972).

A large semantic step or conceptual leap in finding a new topic could be described as a restructuring process and could therefore be identified as an insight (Beeftink, Van Eerde, & Rutte, 2008; Öllinger, Jones, & Knoblich, 2014; Salvi, Bricolo, Kounios, Bowden, & Beeman, 2016). While the order and the number of iterations of these steps can differ between individuals and between impasses in the same conversation, this model assumes that both interlocutors will try to overcome the impasse. However, in other cases the impasses can also mean the end of the conversation, with the interlocutors switching to a ritualized closing remark from step A, C, D, or E.

If conversation is a continuous stream of problem solving, steps C, D and E represent the ill-defined problem of either continuing with the current topic, choosing a new one, or going back to a previous one. In our framework these steps are also processed sequentially, even though the exact order might be different for each problem. In terms of novelty and usefulness, it could be argued that each meaningful unit produced within a topic is non-repetitive. Step D is even more convincing in this regard, since the next speaker is starting a new topic, which is novel by definition - at least on the level of subjective self-discovery. If this topic is appreciated by both speakers, who as a result keep the conversation going, it is useful as well and can therefore be considered creative.

In the following section we will introduce the data that we have used to study impasses and to validate our framework.

METHOD

Recorded telephone conversations offer the advantage of reducing the influence of body language and para-linguistic information in the conversation, as the interlocutors have to express all the information they want to convey via the audio channel only. Plus, the number of interlocutors is limited to two people in classic phone calls. According to our working hypothesis, this increases the chance of observing an impasse.

For these reasons, we analysed some phone calls in the CallHome database, a publicly available corpus of telephone conversations between speakers of American English (Canavan, Graff, & Zipperlen, 1997; MacWhinney, 2007). It was developed by the Linguistic Data Consortium (LDC) in support of the project Large Vocabulary Conversational Speech Recognition (LVCSR) and sponsored by the U.S. Department of Defense. The database consists of 120 unscripted telephone conversations up to 30 minutes in length. The calls were recorded in the 1990s, and in exchange for their participation the speakers received free international phone calls for up to 30 minutes. All the calls originated from North America and most of them were directed to a person, usually a friend or relative, located abroad. The corpus is available at http://www.talkbank.org/CABank/.

We used methods from Conversation Analysis (CA) to study how impasses are solved in spontaneous conversations. Intertwined with Pragmatics, Linguistics and Sociology, CA studies the dynamics of conversation, and tries to explain verbal and nonverbal interactions on the basis of the interlocutors' intentions (Sacks, 1992; Schegloff, 2007). For example, a much discussed topic in CA is the study of turn-taking, i.e. the implicit rules according to which it is appropriate (or not) for an interlocutor to start a new turn (Lerner, 2004; Sacks, Schegloff, & Jefferson, 1974; Stivers et al., 2009). In our model this is partially coded in step A. While an in-depth analysis of conversation dynamics is not required for the current work, we still borrow some techniques and conventions from CA - e.g. for transcribing speech - as a useful set of tools described in Jefferson (2004).

For our analysis we randomly selected 13 out of the 120 available phone calls, with a total length of 245 minutes. In one of these conversations the phone was handed over to a third speaker. We treated this as a separate conversation, which brought the total number of analysed conversations to 14. Two independent raters listened to the phone conversations. They were instructed to *identify impasses (stalemates) in the conversation based on [their] subjective rating*. In addition, they were asked to assign each part of the conversation to either one of the general domains *study*, *work*, or *personal* as well as *identify[ing] meaningful units and write down the general topic*. This scoring technique is very similar to the one used for divergent thinking tasks, as described in Wallach and

Kogan (1965) and Silvia et al. (2008). A change of topic was expected to coincide with an impasse, if interlocutors followed the steps D and E in our model. In addition, we asked our raters to *mark speech interruptions and overlaps*, i.e. when one speaker started talking in the middle of the other person's turn or when both speakers started speaking at the same time after a pause. These overlaps were intended to help us to identify interruptions in step B in our model.

Below we report the transcriptions of three instances of impasses found in the corpus. Following some standard CA conventions, the number in brackets represents silence time (in seconds) and transcriptions enclosed within square brackets indicate overlapping speech. A colon indicates a prolonged segment in speech while transcriptions such as tch, mm, hhh, indicate, respectively, a click of the mouth, an assertive vocalization, and an out-breath. Finally, the left arrow symbols highlight where in the extract our phenomenon of interest, the impasse, is happening. The transcriptions aim at representing what was actually said, without any attempt at interpreting the utterance or correcting grammatical mistakes that might have been made by the speakers.

All the following examples show that the conversation has supposedly passed the sequential stages at B and E and is now in a situation where both interlocutors are experiencing an impasse at the same time. All three cases result in an unnatural interruption of the conversational flow which was identified by the raters as an impasse.

In the first example two friends, who also work in the same sector, are talking about a conference happening in Seattle later in the year, and are discussing whether they will be attending it and in particular the route the flight is going to follow:

1. **S2**: That'd be-'f course Chicago wouldn' t be: (0.5) .hhh too bad either cos
2. it's (0.4) er:
3. **S1**: Polar Air (?)
4. **S2**:(0.4) Yeah (0.3) so it'd be it would be a direct flight
5. **S1**: mmh ←
6. (1.8) ←
7. **S1**: mh[h] ←
8. **S2**: [but] ←
9. (1.8) ←
10. **S1**: well ←
11. (0.6) ←
12. **S2**: We're- hhh. we're g-going home for Christmas

Conversation Analysis 1: CallHome 4112 Conference in Seattle

After a rather long hesitation, made of silences and filled pauses marked as the phenomenon of interest in lines 4 to 10, speaker 2, who also happens to be the last speaker before the impasse, resolves it by completely changing the topic from the conference to the Christmas holidays. This corresponds to the impasse being resolved in step D which in this case also means a change in the domain of the conversation, from work to personal.

In an example from another phone call, two people are talking about how one of them got an evening job in addition to her studies:

1. **S2**: And then I can work (.) like (0.4) evenins during school
2. **S1**: (0.3) A [ha]
3. **S2**: [I don't (.)] have to work in the day
4. (1.8) ←
5. **S1**: tch oh very good ←
6. (1.0) ←
7. **S1**: And how's Lee Roy?

Conversation Analysis 2: CallHome 638 New job

While speaker 2 was the last speaker before the impasse marked in lines 4 to 6, the impasse is resolved by speaker 1. Again, both the topic and the domain are changed, but in addition to the previous example an overlap A between the two speakers occurred (see lines 2 and 3) after which speaker 2 continued to talk as decided in A, but then none of them managed to produce a sentence to continue the topic in C.

Their joint impasse finally seems to be resolved by starting a new topic in D. In this final example, two people are talking about holidays around the 4th July:

1. **S2**: Do you have to work tomorrow?
2. **S1**: (0.3) Yes
3. **S2**: (0.3) Did you have hols yesterday?
4. **S1**: (0.4) Yes
5. **S2**: (0.3) Oh that's good
6. **S1**: (0.3) But and it was paid too
7. **S2**: (0.4) Ohh: is it paid today?
8. **S1**: (0.1) No
9. **S2**: (0.4) Oh ←
10. (1.6) ←
11. **S1**: but er: ←
12. (0.7) ←
13. **S1**: Yeah (0.4) So the gym's going well

Conversation Analysis 3: CallHome 4092 Paid holiday time

In this case, the impasse is resolved by speaker 1, who was also the last one contributing to the topic in line 8. Once again the impasse is overcome by a change in domain, from work to personal, but in this case speaker 1 revisits a topic mentioned earlier in the conversation. This would correspond to step E in Figure 1.

Apart from qualitative data, we extracted some quantitative data relative to the speech signal using the software for phonetic analysis Praat (Boersma & Weenink, 2010). We recorded the total duration of the conversation, the duration of the individual impasses, the identifier of the speaker before and after the impasse, and the time distance between the impasse and the first preceding or following interruption. We also calculated the phonation time of each interlocutor, i.e. the period of time they are speaking without pauses, and the number of syllables. This was obtained counting the nucleus of a syllable by finding voiced peaks in energy surrounded by dips, as described in de Jong and Wempe (2009). Given that each speaker was recorded on a separate audio channel, there is a noticeable difference between the combined phonation time and the sum of the individual times with an overlap of up to 24% of the combined phonation time. The phonation time was used to identify a *dominant* speaker within each conversational pair, i.e. the interlocutor who had the longest phonation time. We also calculated the articulation rate, i.e. the quotient of syllable count over phonation time, which is a measure of how fast a speaker talks (see Jacewicz, Fox, O'Neill, & Salmons, 2009).

RESULTS

Following the instructions *identify impasses (stalemates) in the conversation based on your subjective rating*, rater A identified 99 and rater B 118 impasses in the 14 conversations. After their individual assessments the raters were asked to revisit each of the identified impasses to find a shared agreement on its existence and length. As a result, the raters mutually agreed on a total of 110 impasses - on average about 8 per conversation and one every 2 minutes.

All of these identified impasses are accompanied by silence in the conversation, even though the instructions did not explicitly mention this increase in response latency. This observation strengthens the implied overlap between the term *impasse*, which is commonly used in creativity research, and *gaps* (silences that can be followed by a change of turn) as described in CA (Sacks, Schegloff, & Jefferson, 1974).

We also observed nonverbal vocalizations such as mhh, and interjections such as "well" and "yeah" quite frequently during these impasses, as demonstrated in examples for CallHome 4112 and 4092. This is also consistent with previous findings such as in Park (2010), where the adverb *anyway* is identified as an attempt to move past a conversational impasse and is often preceded or followed by similar vocalizations. More generally this relates to the role of environmental cues to overcome impasses (e.g. Moss, Kotovsky, & Cagan, 2011).

The median length of the 110 identified impasses is *m* = 3.4 *s* with a minimum length of 1.0 *s*. In the cases where the interlocutors switched to a completely new topic after the impasse, the median length of the impasses is *m* = 3.5 *s*; when they resumed a topic that had appeared previously in the conversation, the impasse length is *m* = 3.9 *s*.

In 80 out of the 110 identified impasses (72.7%) the topic changed after the impasse, while in the remaining 30 impasses the interlocutors continued with the same topic as before. While continuing on the same topic does not meet the definition of novel and useful for creative problem solving, the repetitive generation of the same solution can still be perceived as an impasse, as coded by Fleck and Weisberg (2013). Out of the 80 impasses that were followed by a change of topic, 62 (77.5%, 56.4% of all impasses) started a topic novel within the conversation, whereas 18 impasses (22.4%, 16.4% overall) went back to a topic that had been discussed before. While these 62 changes of topic meet the novelty constraint of a creative solution, revisiting a previous topic can also be interpreted as the repetition of a previous solution (see Fleck & Weisberg, 2013).

Looking at the 62 conversations that started a novel topic, 36 of them (58.1%, 32.7% overall) were resolved by the dominant speaker of that conversation, 22 by the non-dominant speaker (35.5%, 20.0% overall), and 4 ended in overlapping speech (6.5%, 3.6% overall). Interestingly, in all these overlaps it was the non-dominant speaker who continued talking afterwards. This is also represented by the fact that the dominant speaker was the last speaker before the impasse in 37 cases (59.7%, 33.6% overall), while, in the remaining 25 cases (40.3%, 22.7% overall), it was the non-dominant speaker. The association between dominance and impasse solving could be related to fluency, which has been shown to be a confounding factor of creativity (Hocevar, 1979; Silvia et al., 2008). However, the fact that dominant speakers spend more time talking means that they have a higher probability of starting a sentence after a gap.

For the 18 impasses that were resolved by revisiting a previously discussed topic, only in 7 cases (38.9%, 6.3% overall) the dominant speaker was talking before, as opposed to the non-dominant speaker, who spoke in 11 cases (61.1%, 10.0% overall). Finally, 11 of these 18 impasses (61.1%, 10.0% overall) were resolved by the dominant speaker, 6 by the non-dominant one (33.4%, 5.4% overall), and 1 (5.6%, 0.9% overall) with an overlap - which was won by the dominant speaker.

In total, 54.5% of the occurrences had a dominant speaker right before the impasse started, and 58.8% were overcome by a dominant speaker.

DISCUSSION

Given the simple instruction of identifying impasses in the conversation corpus, both raters were able to identify a similar number which shows a shared understanding of the task. There was a high overlap between their ratings and they eventually agreed on 110 impasses over 245 minutes of phone calls - which on average translates to about one impasse every two minutes. This confirms our starting argument that the phenomenon which the raters identified as impasses is not only happening inside a conversation, but can also be identified by an observer from the outside. The fact that all the events marked as impasses had an increased response latency additionally supports the idea that the problem solving process can be described using a sequential model (Amabile, 1983; Wallas, 1926). At the same time, this finding ties the observed phenomenon of a *gap* or *lapse* in conversation (Sacks, Schegloff, & Jefferson, 1974) to that of impasse in problem solving (Öllinger, Jones, & Kloblich, 2014; Weisberg, 2015).

Interestingly there was no impasse with a pause of less than 1 second marked by our raters. While the conversations might have had instances of impasses shorter than this, the raters did not identify them. Also, since pauses, nonverbal vocalizations, and interjections are sometimes used as rhetorical figures of speech, our raters might have identified them as such instead of impasses.

Looking at the length of pauses it is worth noting that impasses followed by a new topic are shorter than the ones where the speakers went back to a previous topic. This suggests that during an impasse speakers might look for a solution in the same order across individuals, potentially even in the same order as depicted in Figure 1: first step D and then E. Also, the distribution of impasses that were resolved by a new topic (62), continued on the current one (30), and went back to a previous one (18) supports the idea of a distinct sequence of steps to solve the problem. In this case the continuation on the same topic comes before the revisiting of a previous one which might imply, at least for the case of an impasse, that the order of steps changes from D, C, and E. We did not explore this in more detail as the exact order of steps is not relevant for our argument.

The dominant speakers who are considered to be more active and who spend more time talking, are also more likely to be talking before or after an impasse occurs, thus suggesting that dominant speakers might be more involved in initiating and solving impasses than non-dominant speakers. Again, this supports the correlation between higher fluency and originality in problem solving literature (e.g. Silvia et al., 2008). Interestingly this is not the case for the speakers talking before an impasse which will be resolved with a revisited topic. But this trend might simply be due to random noise in our small sample size. Another explanation could be rooted in individual differences in articulation rate: spending more time talking does not necessarily mean that a speaker will be perceived as more active, since she / he might just be speaking more slowly. The mean articulation rate across all the individual speakers in our sample was A = 4.47 $^{syllables}/_{s}$. While the small sample size does not allow any conclusion based on quantitative analysis for the current paper, it might prove useful to observe the following trend in a future study: non-dominant speakers seem to be talking more often before impasses that are solved without starting a novel topic. This trend is consistent with the idea that speakers with a low fluency fail to create original sentences which will then lead to the observed impasses.

CONCLUSION

We are aware that our results can only be understood as a theoretical exploration since observational studies require a large data set to produce reliable quantitative results. Instead, we use the transcriptions and ratings to maintain the possibility that the observed trends can be interpreted as our framework suggests and as is shown in Figure 1. Our framework is consistent with and supports sequential models of creative problem solving, from social creativity, to neuroimaging studies, and behavioural as well as theoretical work (as a representative sample of work, see Bowden & Jung-Beeman, 2006; Wallach & Kogan, 1965; Wallas, 1926; Weisberg, 2015). For future work we suggest using a larger sample size which would allow, in a quantitative approach, to draw statistical inference and to apply logistic regression techniques to predict, for example, speaker dominance before and after the impasse from the other available factors (topic, domain, duration, articulation rate, etc.).

Sometimes an impasse within a situation might be provoked intentionally by one of the speakers, for example to signal fatigue or to test the other participants' interest in the current topic. Therefore, a future study could be based on qualitative methods to revisit subjective ethnography and compare collected reports with silently delayed responses and occurrences of non-verbal vocalizations during the impasses.

The preliminary results presented in this paper provide evidence that impasses in conversation exist and are perceivable by external listeners. The impasses analysed in our corpus support the hypothesis that impasses in conversation are resolved by a continuous problem solving sequence of search steps in the solution spaces (new topic, previous topic, etc.). They also provide evidence for the conceptual switching hypothesis. These two models are also described as *business-as-usual* and the *insight-sequence* (Newell & Simon, 1972; Ohlsson, 2011). Our framework is neutral regarding which of these models should be used to describe conversations, and it also conforms with the integrated theory proposed in Weisberg (2015). Our suggested framework helps to understand an ongoing conversation as a sequence of time-based events, intertwined with potentially quick, implicit, and distinct steps.

From our reading of the literature as well as the trends observed in this study, we conclude that conversation can be seen as continuous series of creative problem solving. As far as we are aware, this is the first study that links spoken language and creativity through the phenomenon of impasses.

REFERENCES

Amabile, T. M. (1983). The social psychology of creativity: a componential conceptualization. *Journal of Personality and Social Psychology*, *45*(2), 357-376.

Austin, J. L. (1975). *How to do things with words*. Oxford University Press.

Beeftink, F., Van Eerde, W., & Rutte, C. G. (2008). The effect of interruptions and breaks on insight and impasses: do you need a break right now? *Creativity Research Journal*, *20*(4), 358-364.

Boden, M. A. (2004). *The Creative Mind: Myths and Mechanisms* (2nd edition). Routledge.

Boersma, P. & Weenink, D. (2010). Praat: doing Phonetics by computer.

Bowden, E. M. & Jung-Beeman, M. (2006). Methods for investigating the neural components of insight. *Methods*, *42*, 87-99. doi:doi:10.1016/j.ymeth.2006.11.007

Canavan, A., Graff, D., & Zipperlen, G. (1997). Callhome American English speech. *Linguistic Data Consortium*.

Chambliss, C. A. & Feeny, N. (1992). Effects of sex of subject, sex of interrupter, and topic of conversation on the perceptions of interruptions. *Perceptual and Motor Skills*, *75*(3f), 1235-1241.

Clark, H. H. (1996). *Using language*. Cambridge University Press.

Cranford, E. A. & Moss, J. (2012). Is insight always the same? a protocol analysis of insight in compound remote associative problems. *The Journal of Problem Solving*, *4*. doi:10.7771/1932-6246.1129

Cummins, F., Li, C., & Wang, B. (2013). Coupling among speakers during synchronous speaking in English and Mandarin. *Journal of Phonetics*, *41*(6), 432-441.

de Jong, N. H. & Wempe, T. (2009, May). Praat script to detect syllable nuclei and measure speech rate automatically. *Behavior Research Methods*, *41*(2), 385-390. doi:10.3758/BRM.41.2.385

Demuth, C. & Glaveanu, V. P. (2016). Creativity - A New Vocabulary. In V. P. Glaveanu, L. Tanggaard, & C. Wegener (Eds.), (Chap. Language, pp. 52-60). Palgrave Macmillan. doi:10.1057/9781137511805

Donaldson, S. K. (1979). One kind of speech act: How do we know when we're conversing? *Semiotica*, *28*(3-4), 259-300.

Fischer, G., Giaccardi, E., Eden, H., Sugimoto, M., & Ye, Y. (2005). Beyond binary choices: Integrating individual and social creativity. *International Journal of Human Computer Studies*, *63*, 482-512. doi:10.1016/j.ijhcs. 2005.04.014

Fleck, J. I. & Weisberg, R. W. (2013). Insight versus analysis: Evidence for diverse methods in problem solving. *Journal of Cognitive Psychology*, *25*, 436-463. doi:10.1080/20445911.2013.779248

Gilhooly, K. J., Fioratou, E., Anthony, S. H., & Wynn, V. (2007). Divergent thinking: Strategies and executive involvement in generating novel uses for familiar objects. *British Journal of Psychology*, *98*, 611-625. doi:10.1348/096317907X173421

Grice, H. P. (1957). Meaning. *The Philosophical Review*, 377-388.

Grice, H. P. (1975). Syntax and semantics: Speech acts Volume 3. In P. Cole & J. L. Morgan (Eds.), (Chap. Logic and conversation. pp. 41-58). New York: Academic Press.

Guilford, J. P. (1967). *The Nature of Human Intelligence*. McGraw-Hill.

Halliday, M. A. K. (1994). *An Introduction to Functional Grammar*. London: Edward Arnold.

Hocevar, D. (1979). Ideational Fluency as a Confounding Factor in the Measurement of Originality. *Journal of Educational Psychology*, *71*(2), 191-196.

Howard, T., Culley, S., & Dekoninck, E. (2007). Creativity in the Engineering Design Process. In *International conference of engineering design*.

Jacewicz, E., Fox, R. A., O'Neill, C., & Salmons, J. (2009, July). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(02), 233-256. doi:10.1017/s0954394509990093

Jefferson, G. (2004). Conversation analysis: studies from the first generation. In G. H. Lerner (Ed.), (Chap. Glossary of transcript symbols with an introduction, pp. 13-31). John Benjamins Publishing.

Jordan, M. E., Lanham, H. J., Crabtree, B. F., Nutting, P. A., Miller, W. L., Stange, K. C., & McDaniel, R. R. (2009). The role of conversation in health care interventions: Enabling sensemaking and learning. *Implement Sci*, *4*(15), 1-13.

Kaplan, M. A. (1997). Learning to converse in a foreign language: The reception game. *Simulation* & *Gaming*, *28*(2), 149-163.

Kaufman, J. C. & Beghetto, R. A. (2013). Do people recognize the four Cs? Examining layperson conceptions of creativity. *Psychology of Aesthetics, Creativity, and the Arts*, *7*(3), 229-236. doi:10.1037/a0033295

Lerner, G. H. (2004). *Conversation analysis: Studies from the first generation*. John Benjamins Publishing.

MacWhinney, B. (2007). The Talkbank Project. In *Creating and Digitizing Language Corpora* (pp. 163-180). Springer.

McCarthy, M. (2001). The Cambridge Guide to Teaching English to Speakers of Other Languages. In D. Nunan & R. Carter (Eds.), (Chap. Discourse, pp. 48-55). Cambridge: Cambridge University Press.

Moss, J., Kotovsky, K., & Cagan, J. (2011). The effect of incidental hints when problems are suspended before, during, or after an impasse. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *37*(1), 140-148. doi:10.1037/a0021206

Newell, A. & Simon, H. A. (1972). *Human Problem Solving*. Prentice-Hall International, Inc.

Ohlsson, S. (2011). *Deep Learning: How the Mind Overrides Experience*. Cambridge University Press.

Öllinger, M., Jones, G., & Knoblich, G. (2014). The dynamics of search, impasse, and representational change provide a coherent explanation of difficulty in the ninedot problem. *Psychological Research*, *78*(2), 266-275.

Park, I. (2010). Marking an impasse: The use of anyway as a sequence-closing device. *Journal of Pragmatics*, *42*. doi:10.1016/j.pragma.2010.06.002

Reed, S. K. (2015). The Structure of Ill-Structured (and Well-Structured) Problems Revisited. *Educational Psychology Review*, 1-26. doi:10.1007/s10648-0159343-1

Runco, M. A. & Jaeger, G. J. (2012). The Standard Definition of Creativity. *Creativity Research Journal*, *24*(1), 92-96. doi:10.1080/10400419.2012.650092

Sacks, H. (1992). *Lectures on Conversation, Volumes I and II* (G. Jefferson, Ed.). Oxford: Basil Blackwell.

Sacks, H., Schegloff, E. A., & Jefferson, G. (1974). A simplest systematics for the organization of turntaking for conversation. *Language*, 696-735.

Salvi, C., Bricolo, E., Kounios, J., Bowden, E., & Beeman, M. (2016). Insight solutions are correct more often than analytic solutions. *Thinking* & *Reasoning*. doi:10. 1080/13546783.2016.1141798

Schegloff, E. A. (2007). *Sequence organization in interaction: Volume 1: A primer in conversation analysis*. Cambridge University Press.

Searle, J. R. (1969). *Speech acts: An essay in the philosophy of language*. Cambridge University Press.

Selting, M. (1996). On the interplay of syntax and prosody in the constitution of turn-constructional units and turns in conversation. *Pragmatics*, *6*, 371-388.

Silvia, P. J., Winterstein, B. P., Willse, J. T., Barona, C. M., Cram, J. T., Hess, K. I., ... Richard, C. A. (2008). Assessing creativity with divergent thinking tasks: Exploring the reliability and validity of new subjective scoring methods. *Psychology of Aesthetics, Creativity, and the Arts*, *2*(2), 68-85. doi:doi:10.1037/1931-3896. 2.2.68

Simon, H. A. (1973). The Structure of Ill Structured Problems. *Artificial Intelligence*, *4*, 181-201.

Stein, M. I. (1953). Creativity and Culture. *The Journal of Psychology: Interdisciplinary and Applied*, *36*(2), 311-322. doi:10.1080/00223980.1953.9712897

Stivers, T., Enfield, N. J., Brown, P., Englert, C., Hayashi, M., Heinemann, T., ... Yoon, K. -E., et al. (2009). Universals and cultural variation in turn-taking in conversation. *Proceedings of the National Academy of Sciences*, *106*(26), 10587-10592.

Wallach, M. A. & Kogan, N. (1965). *Modes of Thinking in Young Children. A Study of the Creativity-Intelligence distinction*. Holt, Rinehart, and Winston, Inc.

Wallas, G. (1926). *The Art of Thought*. Jonathan Cape London.

Warren, M. (2006). *Features of Naturalness in Conversation*. Pragmatics & Beyond New Series. John Benjamins Publishing. doi:10.1075/pbns.152

Weisberg, R. W. (2015). Toward an integrated theory of insight in problem solving. *Thinking & Reasoning*, *21*(1), 5-39. doi:10.1080/13546783.2014.886625

Corresponding author at: Ilaria Torre, CogNovo, Link3, Plymouth University, Drake Circus, PL4 8AA, Plymouth, UK.
E-mail: Ilaria.Torre@Plymouth.ac.uk

Speech Prosody 2016, May 31 - June 3, 2016, Boston, MA, USA

Behavioural mediation of prosodic cues to implicit judgements of trustworthiness

Ilaria Torre[1], *Laurence White*[1], *Jeremy Goslin*[1]

[1]School of Psychology, Plymouth University, U.K.

ilaria.torre@plymouth.ac.uk

Abstract

Prosodic information is known to play a role in personality attributions, such as judgements of trustworthiness. Research so far has focused on assessing the determinants of such attributions in static contexts, very often in the form of questionnaires, and not much is known about their dynamics, in particular, how direct experience of behaviour over time influences the interpretation of vocal characteristics. We used the investment game, an innovative methodology adapted from game theory studies, to assess how trust attributions – to virtual players acting more or less cooperatively – are affected by the prosodic characteristics of speakers of a range of British English accents. Regression analysis shows that speaker accent, mean pitch, and articulation rate all influence participants investment decisions, our implicit measure of trust. Furthermore, participants interpretations of these prosodic characteristics interact with how the virtual players behave over time. Our findings are discussed with reference to "Size/Frequency Code" and "Effort Code" accounts of prosodic universals.

Index Terms: game theory, trust, accent, mean pitch, articulation rate

1. Introduction

We make personality inferences spontaneously [1, 2, 3], and evidence suggests that the voice of a person just encountered affects the judgements formed almost immediately in the mind of a listener [4]. Here we consider the role of voice characteristics, and prosody in particular, with respect to attributions of trustworthiness, for which there are limited and inconsistent data. In an interview scenario, where participants had to rate the perceived trustworthiness of embodied conversational agents, high f0 early in the interaction predicted low trustworthiness attributions [5]. This finding was consistent with an earlier study based on questionnaire responses, where participants rated speakers with high f0 as "less truthful" [6]. The same study also found that speakers with slow rate of speech were rated as "less truthful" and "more passive" [6]. On the other hand, smiling when speaking (which tends to raise f0 [7]) was found to boost trustworthiness judgements [5]. Finally, accents have long be linked to trustworthiness attributions [8, 9, 10] and were recently shown to affect implicit judgements of trustworthiness, with standard accents being associated with higher trust levels than regional accents [11].

Studies researching specifically the effect of voice characteristics on trustworthiness are scarce, but inferences can be drawn from the literature on the role of voice in deception and in attributions of personality more in general, e.g., the "Big Five" traits [12]. Findings are not wholly consistent, however. Concerning pitch, for example, participants in one deception study consistently raised their f0 when lying [13], while participants in another study *perceived* a lower pitch in the deceptive messages they heard [14]. A further study failed to find any acoustic differences – including in f0 – in the production of deceptive and truthful messages [15]. From a slightly different perspective, two studies found that actors generally had a lower f0 and slower speech rate when acting sarcastic voices than when they were acting sincere voices [16, 17]. Similarly, sincerity in a synthetic voice was associated with greater pitch range and faster articulation rate [18]; slow speech rate, by contrast, was associated with less competence [19]. Finally, higher f0 was associated with higher agreeableness, which also relates indirectly to trust [20].

We are still far from consensus regarding the direction of association between prosodic features – in particular, speech rate and f0 – and trustworthiness. Methodological differences may account, at least in part, for observed discrepancies between studies. Trustworthiness is a complex and dynamic social attribution, difficult to capture fully through explicit questionnaire rating, the predominant method used so far.

We present a study designed to calibrate listeners' dynamically responsive trust attributions to voice characteristics, with a game theory methodology that produces implicit measures of trustworthiness. While questionnaires might give rise to essentially attitudinal responses – e.g. based on social stereotypes linked to accents instead of voice characteristics *per se* – the investment game allows us to study the impact of voice characteristics on trustworthiness without participants realising that this is the focus of the experiment. To our knowledge, no other study has used such an implicit measure to assess the effect of prosody on trust attributions. Our method has the further advantage of taking into account the development of personality attributions with experience: the participants' learning about the other player's behaviour simulates (although in a somewhat contrived manner) the dynamics of personality judgements in real life situations.

2. Method

2.1. The investment game

We used an iterated trust game – "the investment game" – to test how voice-based trust attributions change with experience. Based in game theory [21], it involves making simulated monetary investments to other players, who decide whether to reciprocate the investment or not. The money invested, the dependent variable, can therefore be used as the implicit measure of trust. Experiments using trust games have been carried out to investigate trust attributions to a range of characteristics, including gender [22, 23], race, emotion and reputation [24, 25] and facial expressions [26]. This paper presents a follow-up experiment to [11], in which participants made monetary investments with virtual players associated with different behaviours and accents.

2.2. Participants

Participants were 83 native British English speakers (52 females, 32 males) aged 18-67 (median = 21, SD = 11). They were university undergraduate students who received course credit for participation or members of the public who received monetary compensation. Self-reports on participants' geographical origins were: southwest England (n = 44), southeast England (n = 20), Midlands (n = 7), Wales (n = 5), northwest England (n = 3), East Anglia (n = 2) and northeast England (n = 1) . One English-speaking subject from Canada was eliminated from the data-set, as we needed the participants to have had some contact with a wide range of British English accents.

2.3. Stimuli

The utterances used in the experiment were recorded in a sound-attenuated booth. Research is inconclusive regarding which gender elicits the highest trustworthiness judgement [27], and we eliminated this factor from our experimental design, recording twelve female native British English speakers in their twenties. Three speakers had a Plymouth accent, three a Birmigham accent, three a London accent and three spoke standard southern British English (SSBE). Each speaker read four blocks of 20 sentences (one for each round of the game), all approximately the same length (mean number of syllables per sentence 16.95, SD 1.08). Apart from the first utterance of each block, which served for the virtual player to introduce herself, all other utterances were about strategies to follow in the game, e.g.: "I'm going to return more money now, if you invest more as well"; "The goal of the game is to earn as much money as possible". A noise-removal filter was applied to the recorded utterances, which were then amplitude normalised.

2.4. Procedure

In the investment game, one player has a sum of virtual money which can be invested, in whole or in part, with another player in the hope of receiving an improved return. The amount of money invested is taken as an implicit measure of trust (e.g. [25, 26]). In our version of the game, the participant started with a notional sum of 8 at the beginning of each of the 20 rounds of the game. Each round began with a prompt utterance from the virtual player (see above), and then the participan decided whether to invest all, part, or none of it with the virtual player, who received three times the invested amount. The virtual player was programmed to have one of two behaviours, either returning 120% to 240% of the invested money to the participant (Generous condition) or 0% to 120% (Mean condition), for every round.

Each participant played the investment game four times – twice in the generous condition and twice in the mean condition - with a different virtual player each time. Each virtual player in these four games was associated with a different accent, from one of the 12 pre-recorded voices. The order of behaviours and accents was counterbalanced across participants. The 20 utterances were always in the same position within a block, with random association of utterances to block. Each participant therefore heard all 80 of the prompt sentences, read by four speakers with different accents, across the four game

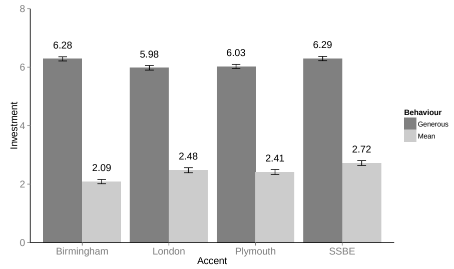


Figure 1: *Mean investments to the four target accents, in the two behaviour conditions*

blocks. Within an accent, speakers were randomly assigned to participants and behaviours in a counterbalanced fashion.

2.5. Prosodic measures

Segmentation and labelling of the individual sound files was done with the MAUS General Web service [28]. The transcriptions thus obtained were then used to extract prosodic measures in Praat [29] and MATLAB, specifically, mean pitch, pitch range, voice quality and articulation rate. Mean pitch was calculated as the mean f0 value for each vowel, then averaged across individual utterances. Pitch range, in order to eliminate potential outliers, was calculated as the difference between the 10th and the 90th percentiles of the mean f0 value for each vowel, as in [30], then averaged across individual utterances. Finally, we used H1-H2 – the difference between the first and second harmonic – as a measure of voice quality, as in [31, 32]. This was calculated using VoiceSauce [33].

3. Results

To determine the effects of game behaviour and vocal characteristics on investments, a mixed-effects regression linear model was fitted to the data using forward stepwise selection, selecting each successive predictor according to the lowest AIC (Akaike Information Criterion [34]). Investment was the dependent variable, behaviour, accent, game turn (i.e. ordinal numbers of the rounds within each game), pitch range, f0, articulation rate and H1-H2 were predictors, and subject was a random factor.

3.1. Effect of game behaviour

The virtual player's behaviour explained the biggest portion of the variance ($\chi^2(4) = 4269.1$, $p < 0.001$). This was expected, since participants learn very quickly to invest higher amounts of money with Generous virtual players than with Mean virtual players. We also found a main effect of game turn ($\chi^2(5) = 96.02$, $p < 0.001$) and a significant interaction between behaviour and game turn ($\chi^2(6) = 191.88$, $p < 0.001$): as previously noted, investments increase in the Generous condition and decrease in the Mean condition as the game progresses.

3.2. Effect of accent

We found an effect of accent, $\chi^2(9) = 30.59$, $p < 0.001$. Post-hoc pairwise comparisons show that average investments to SSBE speakers were higher than the investments to Plymouth ($p = 0.005$), London ($p = 0.022$) and Birmingham ($p < 0.002$)

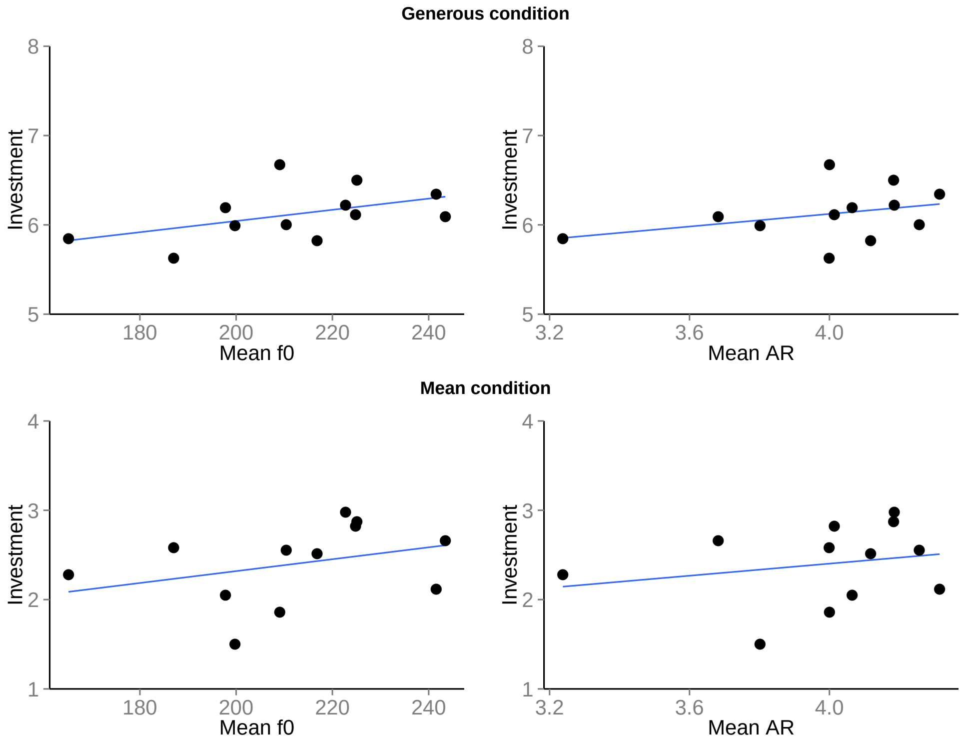


Figure 2: *Scatterplot of mean f0 and articulation rate (AR) against investment in the Generous behaviour condition (top) and Mean behaviour condition (bottom)*

speakers, with no other pairs showing significant differences. There was also an interaction between accent and behaviour, $\chi^2(12) = 13.63$, $p = 0.003$. Figure 1 shows investment according to accent and behaviour. Pairwise comparisons showed that the investments in the Generous condition were significantly higher with SSBE speakers than with speakers from Plymouth (p = 0.04) and London ($p = 0.014$), and that investments to Birmingham speakers were significantly higher than investments to London ($p = 0.017$) and Birmingham speakers ($p = 0.041$). By contrast, in the Mean condition, investments to SSBE, Plymouth and London speakers were all higher than the investments to Birmingham speakers ($p < 0.001$, $p = 0.017$, $p = 0.004$ respectively), and investments to SSBE speakers were also higher than investments to Plymouth speakers ($p = 0.025$).

3.3. Effect of prosody

The regression model showed an effect of f0 ($\chi^2(13) = 5.86$, p = 0.015), while articulation rate approached significance ($\chi^2(15)$ = 3.46, $p = 0.063$). Overall, as shown in Figure 2 , higher pitch and faster articulation rate were associated with higher investments. As we interpret the participants' monetary investments in the game as an implicit measure of trust, these results are largely consistent with previous findings (e.g. [5]). There were no effects of pitch range and H1-H2 on investments.

We found an interaction between f0 and game turn ($\chi^2(14)$ = 12.69, $p < 0.001$), with higher f0 more strongly associated with higher investment earlier in the game. There was an interaction between f0 and articulation rate ($\chi^2(16) = 5.38$, p = 0.02), with greater investment for high f0 at fast rate. There was an interaction between articulation rate and behaviour ($\chi^2(17)$ = 4.58, $p = 0.032$), with the effect of faster rate on boosting investment being somewhat greater in the Mean condition (Figure 2). We discuss the interpretation of these effects of prosody on trustworthiness below.

4. General Discussion

As previously found, the investment game allows us to assess the effect of voices on implicit trust judgements. Standard southern British English attracted higher overall investment than the other British English accents, but – in accordance with our earlier work [11] – there were differential effects of virtual player accents according to their behaviour. In particular, the Birmingham accent attracted relatively high investment in the Generous condition, but the lowest investment in the Mean condition. We are still exploring the underlying reasons for this contrast, which may relate to social stereotypes, individual voice quality or a combination of the two.

Two prosodic features previously associated with trust – high pitch and fast articulation rate – influenced our participants' monetary investments. This is consistent with some, but not all previous studies on attributions of trustworthiness and associated characteristics [5, 14, 16, 17, 18, 19, 20], although the methodology we employed to establish an implicit measure of trust is radically different.

The greater trustworthiness of voices with higher pitch may be related to the "Size/Frequency Code" theory [35], based on higher f0 being generally associated with a smaller larynx and therefore smaller body size. As a consequence, we tend to associate lower f0 with dominance and assertiveness, and higher f0 with friendliness and cooperativeness [36, 37]. In a type of interaction where trust is required, listeners might be more inclined to attribute trustworthiness to a speaker who is perceived as friendly rather than dominating. We note that this interpretation can only be applied to female speakers on the basis of our methodology; it remains to be seen if higher f0 also accords with trustworthiness in male speakers. Furthermore, the relationship between f0 and trust attributions is manifestly not likely to be linear through the range of possible values, partially evidenced in our study by how f0 effects on investment interact with game turn and articulation rate.

The "Effort Code" account [38] may help to explain the association we found between faster articulation rate and higher trustworthiness attributions. Speaking faster can be seen as a corollary of greater conversational effort and hence more engagement in the interaction; manifestly, perceived engagement is likely be seen as a sign of trustworthiness. As with f0, there must be limits to any trust-boosting effects of rate however, and we found differential effect of rate according to whether the virtual player was behaving generously or meanly. It may be considered whether a faster rate impacts on participants' sense of response urgency and therefore on investment decisions. Previous studies have shown that, under conditions of time pressure, a faster rate in spoken prompts may encourage quicker responding [39, 40], although such pressure was not explicitly present in our context. In addition, a fast speech rate of a synthesized voice did not influence the persuasiveness of a commercial message [41]. Naturally, given the complex interactions that emerged among speech rate, accent and our other variables of interest, future studies tailored to unravel them are needed.

There are likely to be voice characteristics contributing to trustworthiness attributions that have not been considered in this study. Furthermore, the effect of the listener's personality and their context-driven expectations should also be examined in a comprehensive account. However, our key findings with respect to prosody concur with most of the foregoing literature: higher f0 and faster rate are associated with higher speaker trustworthiness. Given that we have used an implicit measure with no explicit reference to speaker voice in the experimental procedure, we can be confident that the findings reflect genuine paralinguistic effects rather than stereotype-based attributions.

5. Acknowledgements

This work was supported by CogNovo (FP7-PEOPLE-2013-ITN-604764), a project funded by the EU Marie Curie programme.

6. References

[1] J. S. Uleman, S. A. Saribay, and C. M. Gonzalez, "Spontaneous inferences, implicit impressions, and implicit theories," *Annual Review of Psychology*, vol. 59, pp. 329–360, 2008.

[2] C. Nass, Y. Moon, B. Fogg, B. Reeves, and D. C. Dryer, "Can computer personalities be human personalities?" *International Journal of Human-Computer Studies*, vol. 43, no. 2, pp. 223–239, 1995.

[3] J. Willis and A. Todorov, "First impressions: making up your mind after a 100-ms exposure to a face." *Psychological Science*, vol. 17, no. 7, pp. 592–598, 2006.

[4] P. McAleer, A. Todorov, and P. Belin, "How do you say Hello? Personality impressions from brief novel voices." *PLoS ONE 9(3): e90779*, 2014.

[5] A. C. Elkins and D. C. Derrick, "The sound of trust: voice as a measurement of trust during interactions with embodied conversational agents." *Group Decision and Negotiation*, vol. 22, no. 5, pp. 897–913, 2013.

[6] W. Apple, L. A. Streeter, and R. M. Krauss, "Effects of Pitch and Speech Rate on Personal Attributions," *Journal of Personality and Social Psychology*, vol. 37, no. 5, pp. 715–727, 1979.

[7] V. C. Tartter and D. Braun, "Hearing smiles and frowns in normal and whisper registers," *Journal of the Acoustical Society of America*, vol. 96, no. 4, pp. 2101–2107, 1994.

[8] H. Bishop, N. Coupland, and P. Garrett, "Conceptual accent evaluation: Thirty years of accent prejudice in the UK." *Acta Linguistica Hafniensia: International Journal of Linguistics*, vol. 37, no. 1, pp. 131–154, 2005.

[9] J. A. Dixon, B. Mahoney, and R. Cocks, "Accents of guilt? effects of regional accent, race, and crime type on attributions of guilt," *Journal of Language and Social Psychology*, vol. 21, no. 2, pp. 162–168, 2002.

[10] J. N. Fuertes, W. H. Gottdiener, H. Martin, T. C. Gilbert, and H. Giles, "A meta-analysis of the effects of speakers accents on interpersonal evaluations," *European Journal of Social Psychology*, vol. 42, pp. 120–133, 2012.

[11] I. Torre, J. Goslin, and L. White, "Investing in accents: How does experience mediate trust attributions to different voices?" in *Proceedings of the 18th International Congress of Phonetic Sciences*, Glasgow, U. K., 10-14 August 2015.

[12] L. A. Pervin, *Personality: Theory and Research*. New York: Wiley, 2001.

[13] L. Anolli and R. Ciceri, "The voice of deception: vocal strategies of naive and able liars," *Journal of Nonverbal Behavior*, vol. 21, no. 4, pp. 259–284, 1997.

[14] M. Zuckerman, R. S. DeFrank, J. A. Hall, D. T. Larrance, and R. Rosenthal, "Facial and vocal cues of deception and honesty," *Journal of Experimental Social Psychology*, vol. 15, no. 4, pp. 378–396, 1979.

[15] C. Kirchhbel and D. M. Howard, "Detecting suspicious behaviour using speech: Acoustic correlates of deceptive speech - an exploratory investigation," *Applied Ergonomics*, vol. 44, pp. 694–702, 2013.

[16] H. S. Cheang and M. D. Pell, "The sound of sarcasm," *Speech Communication*, vol. 50, pp. 366–381, 2008.

[17] R. Rao, "Prosodic consequences of sarcasm versus sincerity in Mexican Spanish," *Concentric: Studies in Linguistics*, vol. 39, no. 2, pp. 33–59, November 2013.

[18] J. Trouvain, S. Schmidt, M. Schrder, M. Schmitz, and W. J. Barry, "Modelling personality features by changing prosody in synthetic speech," in *Proceedings of the third International Conference on Speech Prosody*, Dresden, Germany, 2-5 May 2006.

[19] B. Smith, B. Brown, W. Strong, and A. Rencher, "Effects of speech rate on personality perception," *Language and Speech*, vol. 18, pp. 145–152, 1975.

[20] M. Imhof, "Listening to voices and judging people," *The International Journal of Listening*, vol. 24, pp. 19–33, 2010.

[21] W. Poundstone, *Prisoner's Dilemma*. New York: Doubleday, 1992.

[22] A. Boenin and D. Serra, "Gender pairing bias in trustworthiness," *The Journal of Socio-Economics*, vol. 38, pp. 779–789, 2009.

[23] A. Chaudhuri, T. Paichayontvijit, and L. Shen, "Gender differences in trust and trustworthiness: Individuals, single sex and mixed sex groups." *Journal of Economic Psychology*, vol. 34, pp. 181–194, 2013.

[24] M. I. Tortosa, J. Lupiez, and M. Ruz, "Race, emotion and trust: An ERP study." *Brain Research*, vol. 1494, pp. 44–55, 2013.

[25] D. A. Stanley, P. Sokol-Hessner, D. S. Fareri, M. T. Perino, M. R. Delgado, M. R. Banaji, and E. A. Phelps, "Race and reputation: perceived racial group trustworthiness influences the neural correlates of trust decisions," *Philosophical Transactions of the Royal Society B*, vol. 367, pp. 744–753, 2012.

[26] E. Krumhuber, A. S. R. Manstead, D. Cosker, D. Marshall, P. L. Rosin, and A. Kappas, "Facial dynamics as indicators of trustworthiness and cooperative behavior," *Emotion*, vol. 7, no. 4, pp. 730–735, 2007.

[27] C. Nass and S. Brave, *Wired for speech: How voice activates and advances the human-computer relationship.* Cambridge, MA: MIT Press, 2005.

[28] F. Schiel, "Automatic phonetic transcription of non-prompted speech," in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, United States, 1-7 August 1999, pp. 607–610.

[29] P. Boersma and D. Weenink, "Praat: doing phonetics by computer," 2010.

[30] D. Patterson and D. R. Ladd, "Pitch Range Modelling: Linguistic Dimensions of Variation," in *Proceedings of the 14th International Congress of Phonetic Sciences*, San Francisco, United States, 1-7 August 1999, pp. 1169–1172.

[31] K. Johnson, *Acoustic and Auditory Phonetics*, 2nd ed. Oxford: Blackwell, 2002.

[32] M. Garellek, P. Keating, C. M. Esposito, and J. Kreiman, "Voice quality and tone identification in White Hmong," *Journal of the Acoustical Society of America*, vol. 133, pp. 1078–1089, 2013.

[33] Y.-L. Shue, P. Keating, C. Vicenik, and K. Yu, "Voicesauce: A Program for Voice Analysis," in *Proceedings of the 17th International Congress of Phonetic Sciences*, Hong Kong, China, 17-21 August 2011, pp. 1846–1849.

[34] K. Aho, D. Derryberry, and T. Peterson, "Model selection for ecologists: the worldviews of AIC and BIC," *Ecology*, vol. 95, no. 3, pp. 631–636, March 2014, paper reference for AIC selection.

[35] J. J. Ohala, "Cross-language use of pitch: An ethological view," *Phonetica*, vol. 40, pp. 1–18, 1983.

[36] ——, "The acoustic origin of the smile," *Journal of the Acoustical Society of America*, vol. 68, 1980.

[37] J. Hirschberg, "The Pragmatics of Intonational Meaning," in *Proceedings of the first International Conference on Speech Prosody*, Aix-en-Provence, France, 11-13 April 2002.

[38] C. Gussenhoven, "Intonation and interpretation: phonetics and phonology," in *Proceedings of the first International Conference on Speech Prosody*, Aix-en-Provence, France, 11-13 April 2002.

[39] H. Shintel and H. C. Nusbaum, "The sound of motion in spoken language: Visual information conveyed by acoustic properties of speech," *Cognition*, vol. 105, no. 3, pp. 681–690, 2007.

[40] ——, "Moving to the speed of sound: Context modulation of the effect of acoustic properties of speech," *Cognitive science*, vol. 32, no. 6, pp. 1063–1074, 2008.

[41] C. Jones, L. Berry, and C. Stevens, "Synthesized speech intelligibility and persuasion: Speech rate and non-native listeners," *Computer Speech & Language*, vol. 21, no. 4, pp. 641–651, 2007.