DISSERTATION

Psychological and neurophysiological correlates of social vocal control

---

STELLA GULDNER

Inaugural dissertation submitted in partial fulfillment of the requirements for the degree Doctor of Social Sciences in the Graduate School of Economic and Social Sciences at the University of Mannheim

Supervisors:

Prof. Dr. Dr. h.c. Dr. h.c. Herta Flor

Prof. Dr. Carolyn McGettigan

Dean of the School of Social Sciences:

Prof. Dr. Michael Diehl

Thesis Evaluators:

Prof. Dr. Frank Krüger

Prof. Dr. Dagmar Stahlberg

Date of Defense:

15.04.2021

For Elisabeth and Horst Süßdorf whose voices, surely, must bring sunshine to the heavens.

Table of Contents

Abstract

Flexible vocal behaviour is an important paralinguistic signal in social interactions and contributes to how speakers are perceived by listeners. However, it is unclear whether speakers can volitionally modulate their voices in order to express social information about themselves to others. Moreover, it is unclear which speaker characteristics and neural circuits support volitional social voice modulation, despite its implications for understanding social impairments in mental disorders such as autism or psychopathy. Within three research articles, the psychological and neurophysiological mechanisms involved in social vocal control are addressed.

In the first article, the efficacy of volitional voice changes to navigate listener impressions are investigated. Here speakers modulated their voice to express social traits (e.g. sounding attractive, intelligent or likeable). We measured social vocal control by comparing outcome ratings for each expressed trait using a sample of naïve listeners. We found that speakers evoked specific increases in trait perception in listeners in comparison to their neutral voice. These evoked trait ratings were clustered along the two main dimensions of the social trait space, which is a domain-general heuristic to inform social perception and subsequent social behaviour. Thus, supporting the notion that volitional vocal behaviour is a signal that can be volitionally employed by speakers to manage how they are heard by others.

In the second article, we used functional magnetic resonance imaging to describe neural mechanisms supporting social vocal control. The results showed that speakers specifically engage midline and temporal cortical regions to modulate their voice, which are reliably involved in social processing and trait representations. Moreover, we show, that these regions are functionally coupled to a core vocal motor control area in left inferior frontal gyrus during goal-directed social voice change. This suggests the integration of social scripts and motor plans to achieve voluntary control over articulators to evoke specific trait impressions in listeners.

In a third article, we investigated which speaker characteristics contribute to how well speakers can modulate their voice to express social information to listeners, both on the behavioural and neural level. We found that speakers' dispositional empathy, Machiavellistic and psychopathic traits were not associated with general performance in social vocal control, but that speakers with higher Machiavellistic traits performed better in evoking voice-based likeability ratings in listeners. On the neural level, both social

processing, premotor and sensorimotor feedback regions supported the specificity with which speakers could express trait information to listeners.

In summary, this thesis highlights the importance of vocal behaviour in expressing ourselves to others in a way that is in line with a speaker's goals in an interaction and offers implications for understanding social impairments in psychopathologies such as autism spectrum disorders and psychopathy. We suggest that volitional voice modulation is only one of multiple processes contributing to vocal flexibility and that future work is needed to characterize how social vocal control is achieved in, and supports communication in life interactions.

1 Introduction

The voice is often viewed as a stable personality characteristic, but just as facial expressions transcend the face, how we use our voice is a dynamic and purposeful behaviour to express ourselves to others. It contributes to how speakers are perceived and judged by others, and thus, can have important social consequences. While an effective use of the voice can have far-reaching social advantages, inadequate or abnormal vocal behaviour might contribute to the prevalent social difficulties observed in a number of psychological disorders. Although a vast amount of research focuses on how the voice is heard by listeners, much less is known about how the voice is used by speakers and which factors contribute to beneficial vocal modulation. The focus of this cumulative thesis is to study voluntary modulation of the voice to express social information and is based on the following three manuscripts:

Guldner, S., McGettigan, C., Nees, F., Flor, H. (2020). *Navigating interactions through vocal control: Voluntary social trait expression in voices.* Manuscript submitted.

Guldner, S., Nees, F., & McGettigan, C. (2020). Vocomotor and Social Brain Networks Work Together to Express Social Traits in Voices. *Cerebral Cortex*, *30*(11), 6004–6020. https://doi.org/10.1093/cercor/bhaa175

Guldner, S., Nees, F., Flor, H., McGettigan, C. (2020). *What makes a vocal chameleon? Individual differences in social vocal control are reflected in activation of vocal motor and social processing regions.* Manuscript in preparation.

In these manuscripts, the efficacy with which social information is expressed in the voice, the neural mechanisms underlying social vocal control, and the psychological and physiological factors contributing to individual differences in social vocal control are investigated. In the introductory chapter, first the concept of vocal behaviour is introduced and a review of previous work on vocal control and its functional relevance in social interactions is provided, followed by a description of voice anatomy and previous work on the neurophysiological underpinnings of voluntary vocal control and socio-cognitive processing that might be related to social vocal control. Then, the three manuscripts that form the core of the thesis are presented, before the main findings and their implications will be discussed. Lastly, an outlook and future research directions will be given.

1.1 Social vocal control

The voice is inherently social. It is the vehicle for the most prevalent form of social communication, that is language, with speech being an intentional communicative act (Grice, 1975). Through the voice, the speaker can communicate abundant nonverbal information to a listener. Both stable characteristics, such as a speaker's age, gender (Puts, Gaulin, & Verdolini, 2006) or physical characteristics (Krauss, Freyberg, & Morsella, 2002), and psychological states, such as emotions or attitudes (e.g. Laukka, 2007; Sauter, Eisner, Calder, & Scott, 2010; Scherer, 1972) or intentions (Hellbernd & Sammler, 2016) are reliably communicated in a speaker's voice. Vocal information is used extensively by listeners. Even slight changes in vocal behaviour can express subtle emotional or intentional states of a speaker (Karpf, 2007). In fact, the voice is strikingly dynamic - vocal parameters can vary extensively within speakers, for instance during emotional vocal changes, and can even lead to misidentification of familiar speakers (Lavan, Burston, et al., 2019; for review see Lavan, Burton, Scott, & McGettigan, 2019).

Vocal changes as mentioned above refer to variability in supra-segmental speech characteristics of the voice, that is, vocal paralinguistic behaviour, including variation in voice quality (timbre, breathiness), loudness (intensity), and pitch (fundamental frequency, formant frequencies). Such vocal flexibility can be spontaneously or volitionally motivated (reviewed in Lavan, Burton, et al., 2019). Spontaneous voice changes are often automatic and can occur in response to the external, situational context (e.g. Garnier & Henrich, 2014) or an internal context, such as an experienced emotion (Hammerschmidt & Jürgens, 2007; Paulmann, Furnes, Bøkenes, & Cozzolino, 2016; Pisanski, Nowak, & Sorokowski, 2016; Streeter, Macdonald, Apple, Krauss, & Galotti, 1983; Weusthoff, Baucom, & Hahlweg, 2013). Volitional voice changes, henceforth referred to as vocal control, occur consciously and can be controlled by the speaker. Vocal control occurs with an intention to communicate or conceal information about ourselves to others, and can be observed for instance in vocal imitation (Jansen, Gregory, & Brenier, 2001), vocal disguises (Hirson & Duckworth, 1993; Skoog Waller & Eriksson, 2016; for review see Eriksson, Llamas, & Watt, 2010), singing (Welch, Howard, & Nix, 2019) or acting (Morningstar, Dirks, & Huang, 2017; Raphael & Scherer, 1987).

Social vocal control is a specific instance of vocal control, where voice changes are specifically modulated to express socially relevant information to an interlocutor. Social voice changes have been examined more broadly in the literature, particularly during courtship (e.g. Feinberg, Jones, Little, Burt, & Perrett, 2005). Here, voice changes often occur in response to the perceived attractiveness of the listener (Pisanski, Oleszkiewicz, Plachetka, Gmiterek, & Reby, 2018). Less is known about *volitional* voice changes to express social information, i.e. social vocal control. Expressing social information volitionally might be related to cultural display rules. These rules determine when and how expressive behaviours are adequate in a specific culture (for instance wearing black at a funeral) and are learned during ontogeny. Socialization (or enculturation) of the vocal behaviour occurs from an early age on, with children adapting to the vocal patterns associated with stereotypical social roles (Cartei, Cowles, Banerjee, & Reby, 2014; Cartei, Oakhill, Garnham, Banerjee, & Reby, 2020; Redford & Gildersleeve-Neumann, 2009). Plasticity in vocal behaviour continues throughout adult life, too, depending on the various social roles taken over in professional and private settings. Medical doctors, for instance, engage in increasingly loud speech across the course of their training (Kiese-Himmel, Himmel, & Scherer, 2012) and experts decrease voice pitch when giving expert advice (Sorokowski et al., 2019). This highlights, that vocal behaviour is tightly associated with the goal of an interaction and its social consequences. Importantly, speakers can volitionally express both affect (Viscovich et al., 2003) and social emotions (Morningstar et al., 2017), as well as socially relevant character traits in their voice (Hughes, Mogilski, & Harrison, 2014). Hughes et al. (2014) could show that speakers can effectively influence, for instance, how dominant or intelligent they are perceived by volitional voice modulations. However, from their study, it remains unclear how specific, and therefore targeted, social vocal control can be achieved. In summary, the voice is both a highly learned motor skill and a dynamic form of self-expression (Pisanski, Cartei, McGettigan, Raine, & Reby, 2016; Scott & McGettigan, 2016). In fact, it has been described as “the audio version of our personality, our sonic self.” (Karpf, 2007, p.33).

1.2 Significance of vocal control for social interactions

As reviewed above, the voice is a highly flexible, dynamic and social behaviour with functional relevance for social communication by contributing to how speakers are perceived by others. The voice is part of the toolbox humans have in order to adapt to situational context (Locke, 2017) and control over vocal behaviour is necessary to adapt our speech to others. In fact, Burnham et al. (2002) suggest that speakers make intuitive inferences about the emotional and linguistic needs of listeners and adjust their speech parameters to achieve optimal communication accordingly. One of the most prominent examples of this is child-directed speech (Burnham, Kitamura, & Vollmer-Conna, 2002; Dahl, Sherlock, Campos, & Theunissen, 2014; Kitamura & Burnham, 2003), but authors have also described particular speech patterns when speaking to elderly people (Kemper, Finter-Urczyk, Ferrell, Harden, & Billington, 1998) and speech directed at a lover (Farley, Hughes, & LaFayette, 2013).

Moreover, interlocutors also readily converge or diverge in their vocal behaviour (Krauss & Pardo, 2006). Vocal alignment is important to signal and navigate social distance to others (Giles, Coupland, & Coupland, 1991). Vocal convergence is associated with the social context in which it is produced (Pardo, Jay, & Krauss, 2010), occurring often more strongly in direction of the social hierarchy (Gregory & Webster, 1996) or the social desirability of the speaker (Babel, 2012). Importantly, convergence might promote affiliation (Pardo, Gibbons, Suppes, & Krauss, 2012), increase interactional quality (Gregory, Dagan, & Webster, 1997; Gregory, Webster, & Huang, 1993), facilitate communicating emotions (Neumann & Strack, 2000), or even inform of abnormal social functioning if absent (Bone et al., 2014).

Vocal flexibility thus presents an important interpersonal tool, which speakers might use to influence how they are heard and perceived by others. Previous work shows, that even from short, neutral utterances (“Hello”), listeners reliably judge social traits in speakers, such as how likeable (trustworthy), competent (dominant) or intelligent they might be (McAleer, Todorov, & Belin, 2014). The “social voice space” resembles the social trait space described for faces (Oosterhof & Todorov, 2008) and social groups (Fiske, Cuddy, Glick, & Xu, 2002). Trait judgements are rapidly formed (McAleer et al., 2014; Olivola & Todorov, 2010; Todorov, Pakrashi, & Oosterhof, 2009; Willis & Todorov, 2006) and listeners tend to infer additional personality characteristics of speakers based on the perceived traits (Berry, 1990; Hughes & Miller, 2016; Montepare & Zebrowitz-McArthur, 1987; Zuckerman & Driver, 1989). For instance, speakers with younger sounding voices, are also judged to be less competent and assertive (Montepare & Zebrowitz-McArthur, 1987), and attractive voices are automatically associated with attractive faces (Hughes & Miller, 2016). Perhaps not surprisingly, social judgements thus bias subsequent behaviour and attitudes towards a speaker (Chebat, El Hedhli, Gélinas-Chebat, & Boivin, 2007; Elbert & Dijkstra, 2014; Feinberg et al., 2005; Hecht & LaFrance, 1995; Pavela Banai, Banai, & Bovan, 2017; Schroeder & Epley, 2015; Tigue, Borak, O’Connor, Schandl, & Feinberg, 2012). Judgements based on vocal behaviour influence chances in a job interview (Schroeder & Epley, 2015), election outcomes (Pavela Banai et al., 2017; Tigue et al., 2012) or interpersonal trust (Montano, Tigue, Isenstein, Barclay, & Feinberg, 2017; O'Connor & Barclay, 2017; Torre, Goslin, White, & Zanatto, 2018). Trait judgements are independent of the familiarity of the speaker (Lavan, Mileva, & McGettigan, 2021) and remain stable across linguistic backgrounds (Baus, McAleer, Marcoux, Belin, & Costa, 2019) and stimulus content (Mahrholz, Belin, & McAleer, 2018). Here, we ask whether and to what degree speakers can influence how they are perceived by listeners to navigate social interactions.

1.3 Individual differences in social vocal control

Meaningful social interactions and connections to others are a pivotal part of human life. Many mental disorders are associated with difficulties in social communication (conceptualized in the research domain criteria; RDoC; Insel et al., 2010). The voice is a central channel for social communication, and characteristic changes in vocal behaviour can be observed in a number of mental disorders (Cohen & Elvevåg, 2014), underlining the importance of understanding the factors contributing to functional social vocal behaviour in interactions.

Early theories propose that metacognition about self and others, such as mental state attributions, are necessary for communication to occur successfully (see also Frith, 2012). This creates a shared intentionality (Grice, 1957) based on the mental representations of others (Sperber, 2000). In fact, previous studies on the perception of socio-affective information in voices indicate that an individual's level of social reactivity is associated with the ability to accurately identify and categorize vocal cues (Amenta, Noël, Verbanck, & Campanella, 2013; Aziz-Zadeh, Sheng, & Gheytanchi, 2010; Jacob, Kreifelts, Nizielski, Schütz, & Wildgruber, 2016; Neves, Cordeiro, Scott, Castro, & Lima, 2018). Clinical work suggests that deficits in social reactivity are associated with changed spontaneous vocal behaviour, for instance in autism spectrum disorder (Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2017) or psychopathy (Louth, Williamson, Alpert, Pouget, & Hare, 1998). Social reactivity encompasses both cognitive and affective empathy components (Cuff, Brown, Taylor, & Howat, 2014). While cognitive empathy describes an individual's ability to mentalize about another person, understand their intentions, beliefs and thoughts and take over their perspective, affective empathy describes the ability to resonate with another person emotionally (Cuff et al., 2014; Reniers, Corcoran, Drake, Shryane, & Völlm, 2011). In fact, cognitive empathy is thought to be under voluntary control, whereas affective empathy is automatic (Bird & Viding, 2014). We therefore suggest that cognitive empathy skills might be specifically important to volitionally achieve functional social voice modulations, i.e. social vocal control ability.

Moreover, although clinical psychopathy is associated with abnormal spontaneous vocal behaviour, this might be different for volitional social voice changes. Given that the voice might be an important channel for beneficial self-presentation and managing social interactions, traits associated with high social opportunism and manipulation might be related to increased ability for social vocal control. These traits encompass both psychopathic and Machiavellisitc traits and together with narcissism, are subsumed under "dark personality traits" in the general population (Paulhus & Williams, 2002). Specifically, psychopathic and Machiavellisitc traits are closely related to high levels of manipulation and social exploitation, strategically so in Machiavellianism (Jones & Paulhus, 2009; Paulhus & Williams, 2002). The voice has not been previously studied as a potential and volitional tool to influence others, but previous work shows an interesting versatility to communicate volitional socio-affective information nonverbally. For instance, a sample of inmates with clinical psychopathy performed better in volitional emotional facial mimicry, leading to more accurate emotion recognition in naïve raters, as well as higher ratings of authenticity in comparison to a non-psychopathic control group. Importantly, this was in the face of a non-existent congruent affective state in the psychopathic sample (Book et al., 2015), suggesting skilled voluntary nonverbal social signalling. Moreover, Machiavellianism is associated with strategically used speech convergence in social interactions, specifically when it is profitable for the speaker (Muir, Joinson, Cotterill, & Dewdney, 2016). Thus, social vocal control might be a channel through which people with high Machiavellistic and psychopathic tendencies manage their impressions on others.

1.4 Neurophysiological mechanisms of vocal control

Voice physiology. Phonation has traditionally been described along lines of the source-filter model of speech (Fant, 1960). This model is an approximation of the mechanisms driving voice production, specifically by respiratory energy forcing air from the lungs through the trachea, glottis and larynx, where the passing air stimulates the vocal folds into vibration. The sound waves produced by the vocal folds are then modulated in the vocal tract, through manipulation of its length and shape, and radiated from the mouth. The percept of a voice can be described as the result of three properties: fundamental frequency (pitch), intensity (loudness) and voice quality (roughness/harmonics). All of these are the result of the anatomical properties of the larynx and supra-laryngeal vocal tract, representing the stable aspects of an individual's voice. Within these anatomical limitations, however, temporary modulation can be achieved in response to social or emotional contexts (e.g. Hammerschmidt & Jürgens, 2007; Pisanski, Cartei, McGettigan, Raine, & Reby, 2016) or during emotion expression (Lee, Bresch, Adams, Kazemzadeh, & Narayanan, 2006).

Vocomotor network. In order to speak, a vast amount of orofacial, laryngeal and respiratory muscles work in accordance with each other as an outcome of extremely rapid motor commands required for vocal control. While the primary sensorimotor and premotor cortical representations of the vocal tract and articulators allow to control facial, tongue and pharyngeal muscles, the laryngeal muscle groups map onto the laryngeal motor cortex. This area has been proposed to be involved in fine-tuned laryngeal control for speech (Simonyan & Horwitz, 2011), as well as volitional control over paralinguistic acoustic features of voice, for instance pitch (Dichter, Breshears, Leonard, & Chang, 2018). Voice production encompasses feedforward and feedback operations to achieve dynamic adaptation of vocal behaviour to a target output. Computational models of speech production suggest that particularly left inferior frontal gyrus (IFG) houses sound maps for speech, which are then used for motor command initiation and error monitoring in somatosensory (supramarginal gyrus; SMG) and auditory areas (primary auditory cortex; A1, and superior temporal gyrus; STG; Tourville & Guenther, 2011). Voluntary control over this system is achieved over practice during childhood and ultimately allows the acquisition and production of verbal signals (Tourville & Guenther, 2011). Recent work has established a core set of regions involved in volitional control over vocal modulations. This vocomotor network includes triangular and opercular parts of IFG, supplementary motor area (SMA), SMG, insula, STG, anterior cingulate cortex (ACC), the basal ganglia and cerebellum (reviewed by Carey & McGettigan, 2016; Pisanski, Cartei, et al., 2016).

Concerning paralinguistic voice changes, previous work has focused on affective speech modulation (e.g. reflected in prosody) or affective nonverbal vocalizations (e.g. laughter, crying). The dual pathway model (Owren, Amoss, & Rendall, 2011) suggests that a fine-tuned sensorimotor cortical control system is involved in articulatory speech processes, whereas a deeper, limbic system is involved in socio-emotional modulation of non-verbal vocalizations (Ackermann et al., 2014; Jürgens, 2009, 2002; Wild, Rodden, Grodd, & Ruch, 2003). Both streams seem to interact to produce the meaningful variation in affective and motivational modulation of the voice during speech production. The anterior cingulate cortex (ACC) has thus been proposed to be a possible hub of integration of emotional and motivational vocal production as well as fine-tuned motor planning for speech, possibly by its role in controlling pitch contour (Belyk & Brown, 2016). In fact, lesions in the ACC are associated with speech lacking emotional intonation (Jürgens & von Cramon, 1982). The ACC shows afferent connections from laryngeal motor cortex and efferent connections to brain stem regions associated with gating and generative functions over the vocal apparatus during emotional vocalizations. The periaqueductal grey (PAG) in the brain stem is in turn directly connected to nucleus ambiguous and its laryngeal motor neurons (Jürgens & Pratt, 1979; Jürgens & Zwirner, 1996; see Figure 1).

Involvement of social processing areas in vocal control. Relatively few functional imaging studies have probed the neural correlates of voluntary paralinguistic expression in the voice during speech. Convergent findings suggest an involvement of cortical and subcortical areas in emotional vocal control (in line with the dual-pathway theory). Cortical areas including middle and posterior STG, bilateral IFG, and ACC and anterior insula. Subcortical areas include involvement of the basal ganglia (BG), Hippocampus (HC) and Amygdala (Aziz-Zadeh et al., 2010; Barrett, Pike, & Paus, 2004; Dogil et al., 2002; Frühholz, Klaas, Patel, & Grandjean, 2015; Klasen et al., 2018; Petri Laukka, Åhs, Furmark, & Fredrikson, 2011; Pichon & Kell, 2013). Of these regions, particularly right posterior STS regions were modulated by the social context in which a speaker volitionally expressed affect in their voice, leading the authors to conclude that STS might be central to integrating social information into computations of voluntary affective modulations (Klasen et al., 2018).

In contrast to the centrality of the voice in social interactions, only a small body of research has investigated neural underpinnings of specifically social vocal behaviour. For instance, McGettigan et al. (McGettigan et al., 2013) showed that during vocal impersonations left IFG, anterior insula and STG were involved in general voice change, but particularly bilateral STS, and a region in posterior cingulate cortex venturing into precuneus, were activated during social impersonations. Moreover, STS subregions were functionally connected to a vast set of regions involved in motor control (e.g. IFG, SMA, cerebellum, somatosensory cortex). In an online acting task in which actors covertly answered questions

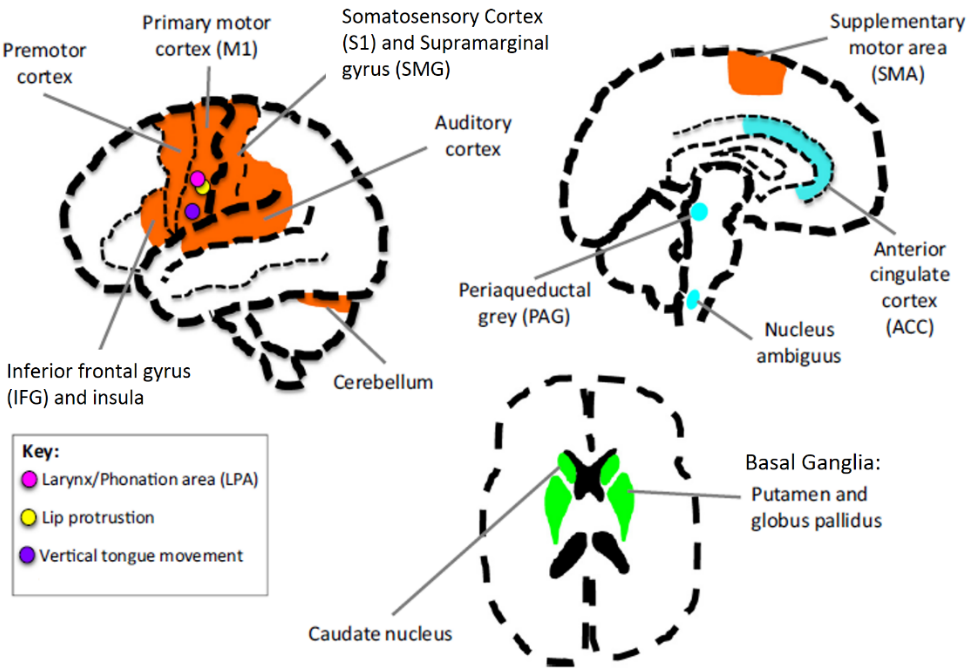


Figure 1. Areas involved in vocal control. Cortical areas are implicated in articulatory motor control for speech (orange), whereas the dual-pathway model suggests that deeper structures, including midline ACC and brain stem nuclei support motivational and affective vocalizations (blue), together with motor and motivational processing in basal ganglia (green). Modified from Pisanski et al. (2016), reprinted without permission.

in the scanner while in role, specifically precuneus and medial prefrontal cortex regions were engaged (Brown, Cockett, & Yuan, 2019).

Together with the temporo-parietal junction (TPJ), this set of regions (STS, mPFC, precuneus) encompasses the social brain network (SBN; Frith & Frith, 2003, for meta-analyses see Bzdok et al., 2013; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014; Van Overwalle, 2009). Activation in SBN is associated with social cognitive processes important for cognitive empathy (such as ToM and mentalizing; Frith & Frith, 2003), for instance, social trait representation (in line with the social trait space), self-referential processing, and future mental state simulations of the self and others (Cabeza, Ciaramelli, & Moscovitch, 2012; de Guzman, Bird, Banissy, & Catmur, 2016; Krueger, Barbey, & Grafman, 2009; Lieberman, Straccia, Meyer, Du, & Tan, 2019; Ma, Vandekerckhove, Van Hoeck, & Van Overwalle, 2012). SBN regions are furthermore associated with inferring social information from voices (Hellbernd & Sammler, 2018; Jiang, Sanford, & Pell, 2017). Similar to the integration of affective and vocomotor streams during affective vocal control, social vocal control might rely on the integration of social processing areas and vocomotor networks.

One previous study has explicitly tested individual differences in performance of affective voice modulation, and found that performance was positively associated with activity in premotor cortex and right IFG, although the right lateralized activations might be due to the affective connotation of the produced voice. This suggests, in line with previous work on motor expertise (Calvo-Merino, Glaser, Grèzes, Passingham, & Haggard, 2005; Krishnan et al., 2018), that premotor regions might be important also to achieve volitional social voice modulation, but this has not been tested so far. Alternatively, as we have seen from clinical evidence, a speaker's capacity to mentalize about listeners might contribute to their ability to express social information in the voice. In favour of this, Aziz-Zadeh et al. (2010) report an association between functional activation during perception of emotional prosody in left IFG with trait empathy, as well as in STS, TPJ and anterior insula with trait psychopathy. In fact, a listener's dispositional empathy modulated activation during social voice processing (inferring speaker confidence) in superior parietal regions and basal ganglia, as well as the functional connectivity between these regions (Jiang et al., 2017). However, this was not tested for volitional emotional voice *production*. Consequently, the underlying socio-cognitive processing strategies that might serve to inform vocomotor plans remain elusive, as does the question of whether motor processing or social processing might contribute to individual differences in social vocal control ability. If social processing contributes to individual differences in social vocal control ability, it would imply a route by which cognitive empathy deficits and social impairments in ASD might contribute to abnormal paralinguistic vocal behaviour.

1.5 Aims and Hypotheses

Given the social consequences of vocal behaviour, there is a striking gap in the literature concerning volitional vocal behaviour in social interactions. The overarching aim of this thesis was to elucidate the psychological and neurophysiological mechanisms involved in social vocal control. To this end, a vocal control task is employed in which participants speak varying exemplars either in their normal voice or while modulating the voice to express specific social traits (e.g. likability, intelligence). As a control condition, speakers modulate their voice to express physical characteristics (e.g. sounding larger or smaller). The vocal recordings are then rated in comparison to the neutral voice by naïve listeners to estimate the efficacy of vocal control.

The goal of study one was to provide evidence of the efficacy of social voice modulations and examine whether speakers can modulate and evoke specific trait perceptions in listeners by volitionally modulating their voice. Here, it was also investigated whether performance in social vocal control ability is predicted by more general vocal control ability (voice control to express physical attributes). Forty healthy volunteers completed the vocal control task. We obtained naïve listener ratings on all social trait scales on all recordings. Based on these multivariate ratings, the following hypotheses were investigated:

1.1 social vocal modulations were expected to evoke an increase in the listeners' perception of the social trait specifically for the expressed trait, beyond the speakers' neutral voice
1.2 vocal modulations were expected to evoke amplifications of trait percepts relative to the speakers' neutral voices that could be explained by the primary axes of the social voice space encompassing affiliation and competence
1.3 Lastly, vocal control performance to express physical attributes (body size) was expected to be positively associated with social vocal control performance.

The second study explored changes in neural activation in response to social vocal control. Twenty-four healthy speakers completed an adapted version of the vocal control task while undergoing an fMRI protocol. Based on the naïve listener ratings of the vocal recordings, we predicted

2.1 significant up-regulation of neural activation during vocal modulation in the VMN (including the pars triangularis and pars opercularis of the left IFG as a central vocal control area, as well as SMA, BG, cerebellum, and insula)
2.2 a higher engagement of areas involved in domain-general social trait and self-referential trait processing, such as pSTS/TPJ and mPFC, and
2.3 that the core vocal control network (left IFG) would be functionally connected to social processing areas during online social voice modulation.

The third study was dedicated to investigating individual differences in social vocal control ability, based on traits related to social reactivity (i.e. trait cognitive and affective empathy, psychopathy and Machiavellianism). To this end, the association between social vocal control ability, social reactivity and functional neural processing during the vocal control task were investigated. Here, the data from the second study was re-analysed together with self-report questionnaire measures of social reactivity, psychopathy and Machiavellianism. The hypotheses were:

3.1 Social vocal control ability is positively associated with dispositional empathy (particularly cognitive empathy) and socially-opportunistic personality traits (Machiavellianism and Psychopathy)

3.2 Social vocal control ability is positively related to increased activation in SBN during social voice production, particularly in mPFC and precuneus.

Additionally, we conducted exploratory analyses to investigate whether

3.3 individual differences in social reactivity are reflected in SBN activity during social voice production, and

3.4 individual differences in social reactivity are reflected in the amount of functional connectivity between l-IFG and SBN during social voice production.

2 Empirical Studies

2.1 Study 1: Navigating interactions through vocal control: Voluntary social trait expression in voices[1]

[1] Guldner, S., McGettigan, C., Nees, F., Flor, H. (2020). *Navigating interactions through vocal control: Voluntary social trait expression in voices.* Manuscript submitted.

Navigating interactions through vocal control: Voluntary social trait expression in voices.

Stella Guldner[1,3*], Carolyn McGettigan[2], Frauke Nees[1,4] and Herta Flor[1]

[1] Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2] Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK

[3] Graduate School of Social and Economic Science, University of Mannheim, Mannheim, Germany

[4] Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany

Corresponding author: Stella Guldner, Central Institute of Mental Health, J5, 63159 Mannheim, email: stella.guldner@zi-mannheim.de, phone: +49 621 1703 6314

Abstract

The voice is a variable and dynamic social behaviour with functional relevance for self-presentation, for example, during a job interview or courtship. Little is known of how the voice can be volitionally modulated to evoke favourable social trait judgements from listeners. In this study, we investigated the mechanisms and efficacy of voluntary social voice modulation. We recorded 40 healthy adult participants during vocal expressions of six social traits (e.g. likeability, dominance), body-size (e.g. sounding larger) or while speaking in a neutral voice. 40 naïve listeners rated all voices on all six trait scales. Speakers' vocal modulations evoked specific exaggerations of trait precepts in naïve-listeners in comparison to their neutral voice, particularly during hostile, likeable, attractive or dominant voice changes. Moreover, trait ratings evoked by social voice modulations could be explained by two principal components relating to perceived affiliation and competence over and above the speakers' neutral voices. The efficacy of vocal control, as the evoked change in ratings on a given trait in comparison to the neutral voice, was unrelated to basic vocomotor control (modulation of body-size). These findings advance our understanding of the mechanisms underlying non-verbal vocal behaviour for social communication and its functional importance for favourable self-presentation.

Keywords: Voice production, social vocal control, self-presentation, social interactions

Navigating interactions through vocal control: Voluntary social trait expression in voices

The voice is at the basis of navigating interactions with others. Beyond language, the voice can be understood as a dynamic form of self-expression [1]: Vocal cues are actively used by listeners to judge a speaker's personality, emotions, or physical attributes [2–7] and, importantly, their communicative intentions [8]. As such, the voice is an important avenue for speakers to manage social interactions, for example, to express positive information about themselves to others. However, it is less clear to what extent the voice can be strategically used in interactions to benefit personal social goals.

Apart from explicit information that is expressed through language, implicit information is also encoded in how we use our voice to convey a message. Vocal modulation can be *functional* in a social interaction. It often responds to the emotional needs of the listener [9,10] and the intention behind speech, be it to educate a child [11] or to engage a potential romantic partner [12–16]. In vocal recordings of naturalistic speed-dates, Pisanski et al. [12] found that vocal modulations occurred in response to individual and group level estimates of a partner's attractiveness. Moreover, when giving expert advice, university faculty members lowered their voice considerably, which in turn was associated with higher ratings of competence and authority in listeners [17]. These findings show that vocal modulation occurs spontaneously and in a goal-directed manner in response to the social context.

Vocal information contributes significantly to the impressions we make on others. Even short, neutral utterances are used by listeners to make rapid social judgements about a speaker [18]. McAleer et al. [18] recorded speakers reading neutral sentences. For each of the speakers they isolated a short utterance containing the word "hello" and asked naïve listeners to rate the voice on a number of socially relevant traits. Principal component analyses showed that the rated traits for each speaker could be described by two primary dimensions corresponding to perceived affiliation (also referred to as warmth or trustworthiness) and competence (also referred to as dominance). In fact, face judgements [19] and social group judgements [20] can also be summarized by how warm/trustworthy or competent we perceive others to be, thereby guiding future interactions and decisions [21,22]. Comparable trait inference has been shown across different languages [23] and speech content [24].

Although rapidly formed impressions do not necessarily reflect reality, listeners tend to agree in their social perception of voices [18] and, moreover, these impressions predict future behaviour towards the speaker. Trait judgements from voices are predictive of success in a job interview [25], election outcomes [26,27], or preferences during courtship [28]. Moreover, once formed, these judgements are often generalized to other personality characteristics [29–31], underlining the social significance of using the voice to maximize the possibility for favourable trait judgements from listeners. Yet, we know very little of how the voice is strategically used in social interactions.

Volitional vocal modulations have long been neglected in voice production research [32]. Such modulations can effectively influence the perception of physical characteristics of a speaker, such as perceived age [33] and emotional or mental states [34]. Moreover, voice modulation can influence the perception of a speaker's personality traits in listeners [35], for example, by modulating their voice to sound more dominant. However, how effectively vocal modulation can be utilized in social interactions, depends on how specifically they lead to a certain trait percept. In their study, Hughes et al. [35] showed that for most expressed traits, naïve listeners perceived the trait to be stronger when it was volitionally expressed than when the talker produced their "neutral" voice. However, the comparison obtained here was only between the neutral voice and the intended trait, and not all other traits. To determine the efficacy with which vocal modulation leads to a specific trait percept, multivariate ratings of the voices (on all possible traits for all voice modulations) are needed. Further, this would allow for an investigation of whether vocal modulations can influence the representation of a speaker's voice in the social voice space - that is, whether a speaker can volitionally amplify certain traits in their voice to influence how they would spontaneously be perceived in terms of their affiliation and competence. Such intentional vocal modulations (i.e. vocal control) becomes particularly important when we care about how we are heard by others in order to achieve personal goals in an interaction, for example getting hired, exerting authority, or attracting a partner. In these instances, sounding competent, dominant, or attractive, respectively, will increase the chances of a successful social transaction. We suggest that speakers are indeed able to exploit knowledge of social trait perception by modulating their voice to achieve desirable social outcomes.

The present study. The goal of this study was to analyse the volitional use of the voice in social interactions. Based on previous findings, the present study aimed to address three open questions in regard to this goal.

First, in order for vocal modulations to be effective, they have to be recognizable by listeners. Therefore, we tested both the sensitivity and specificity with which social traits can be encoded in the voice, that is, whether vocal modulations (in comparison to a speaker's neutral voice) would induce an increase in the trait perception specifically for the intended expressed trait. Second, given the extensive literature showing that social information is intuitively clustered along the dimensions of a social trait space, namely affiliation and competence, we investigated whether volitional vocal modulations are instrumental in the sense that they exploit knowledge of social outcomes (i.e. the perceived trait space). To do this, we measured whether modulated voices would evoke changes in percepts (over and above the perception of the talker's neutral voice) that could be explained by the two primary dimensions of the social voice space. Lastly, we investigated how basic vocal control and social vocal control interact. Is good basic vocal control necessary to express more complex social information? Here, we asked speakers to additionally express a smaller or larger body size, which was again evaluated by naïve listeners. Using these data, we explored whether the performance in basic vocal control would predict performance in social vocal control.

Methods

Participants. Forty native German speakers (age M=22.08, SD=.27, 13 male) participated in this experiment. Participants were recruited as a subgroup of a larger cohort from the European IMAGEN study (for detailed description see Schuman et al., 2010). All participants had normal or corrected-to-normal vision and normal hearing. All participants provided informed written consent prior to their participation and were reimbursed with a monetary reward for their participation, equivalent to an hourly rate of 10€. This study was approved by the research ethics committee the Medical Faculty Mannheim of Heidelberg University (2007-024-N-MA).

Vocal control paradigm. All speakers were asked to complete two vocal control tasks, a social vocal control task and a basic vocal control task. The social vocal control task included six social trait conditions: attractiveness, likeability, hostility, intelligence, confidence and dominance. Basic voice modulation conditions required participants to perform comparable vocal modulations on the vocal tract [36] without an immediate social connotation. These conditions included conveying body-size through the voice (sounding physically smaller or larger; for instructions on all social and basic conditions see supplementary materials S1). Within the basic vocal control task, we additionally recorded non-modulated (neutral) voices as a baseline condition. The exemplars used in all conditions consisted of three neutral German sentences (English translations: "*There are many bridges in Paris*"/"*Many flowers* *bloom in July"*/"*Bears eat a lot of honey"*), which were paired with each vocal control condition.

Design and procedure. The experiment consisted of two vocal control tasks. Before each task, speakers were instructed on the different traits and trial structure. Speakers first performed the basic vocal control task, followed by the social vocal control task. Within each trial of the vocal control tasks, participants saw the target trait and the exemplar sentence on the screen, together with a cue informing them when to speak. At the beginning of each trial, participants had two seconds to prepare. Then the speech cue changed its color to green, marking the start of the recording and instructing the participants to speak the exemplar while expressing the trait in their voice. After four seconds, the speech cue turned back to red indicating the end of the recording. Participants were then presented with a visual analogue scale to rate how strongly they sounded like the target trait in the preceding trial. The response scale ranged from "not at all [target]" (=0) to "very [target]" (=100). Each exemplar and condition combination was repeated in four subsequent trials, forming a trait x exemplar mini-block. The beginning of a new mini-block was indicated by a fixation cross (~one second; see Figure 1).

Participants completed 36 trials (9 mini-blocks of 4 trials each) in the basic vocal control task followed by 72 trials in the social vocal control task (18 mini-blocks of 4 trials each). Together, the first and second part of the experiment took approximately 30 minutes to complete. Visual cues were presented using the Psychophysics toolbox [37,38] in Matlab (version 2016a, the Mathworks, Natick, MA). Recording were obtained on a RØDE NT1-A 1 cardioid condenser microphone (Silverwater, Sydney, Australia) in an anechoic-chamber.

Based on the performance ratings obtained after each recording, we selected one representative recording per participant for each trait condition that had received the participant's own maximal intensity rating for that trait. This was done to ensure that independent evaluations of performance would be based on sounds where the speakers had felt most confident in their expression of the target trait and to alleviate task training effects. In cases where multiple recordings had the same maximal rating, we selected one recording at random.

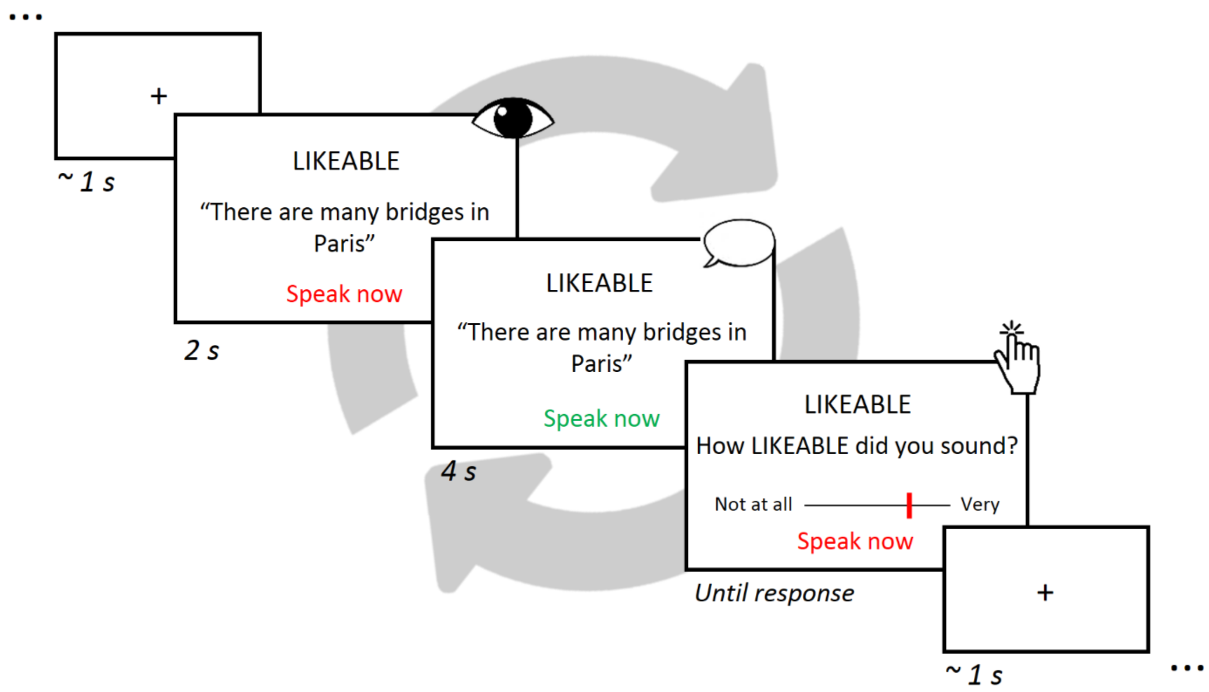


Figure 1. Exemplary trial structure of the vocal control paradigm for likeable voice modulations. A trial consisted of a two second preparation time where the exemplar and modulation condition was presented followed by a four second recording window. After the recording, participants rated their own performance on a visual analogue scale. Each trait and exemplar combination was repeated 4 times in sequence (mini block; grey arrow) to control for training effects. Each mini block was followed by a ~1 s inter-trial interval.

Naïve ratings. Forty naïve listeners (age M=25.38, SD=5.20, 9 male; 39 native German speakers, 1 Polish with German C2 level) were presented with the selected voice recordings obtained in the vocal control tasks to estimate the efficacy of social vocal control for each speaker. Naïve listeners were recruited via the subject databases of University of Mannheim and the Central Institute of Mental Health. Raters received a monetary reward of 10 € per hour or an equivalent in student credit points for their participation and gave their full informed written consent prior to participation.

In the main experiment, each listener heard a subset of 10 speakers. We counterbalanced the subset of speakers to ensure each speaker was heard and rated by at least 10 different raters [35,39]. For each speaker, one recording of each social trait and one recording of the neutral voice was included in the task. These recordings were presented in randomized order to listeners in six rating blocks, where each block was dedicated to rating how much a given voice expressed one of the social target traits. Block order was also randomized. Listeners rated each recording on all traits on 7-point Likert-scales measuring the intensity of expression of the rated trait in the voice from not at all (=1) to very much (=7). To do this, they received the same trait instructions as the speakers.

To assess basic vocal control, we additionally presented listeners with the body-size modulation recordings of each speaker. Here listeners heard two recordings of the same speaker in sequence. One of the recordings was the neutral voice recording of the speaker, whereas the other one was a smaller or larger voice recording. Whether the neutral or modulated voice recording was heard first, was randomized. The listeners were then asked to decide which of the two voices they heard sounded smaller or larger. The response was classified as correct if the rater chose the modulated voice recording in accordance with the modulation task (larger or smaller). Stimuli were presented using the Psychophysics Toolbox [37,38] in Matlab (version 2016a, the Mathwork, Natick, MA) on Headphones (Sennheiser, Wedemark-Wennebostel, Germany) in an anechoic chamber.

Data analyses

Data analysis was conducted in R (http://www.R-project.org/). To assess the effect of social vocal control, we calculated the average change in naïve ratings as a function of the neutral versus the modulated voice samples, individually for each speaker and each trait. We subtracted the mean trait ratings for the neutral voice recordings on a given trait scale from the mean ratings obtained on this trait scale for each modulated voice sample, per speaker. To illustrate this with an example: for intelligence, we separately subtracted the mean "intelligent" ratings for the neutral voice from the mean "intelligent" ratings for the attractive, likeable, hostile, intelligent, confident and dominant voice. Thus, we obtained the change in mean ratings as a function of each expressed trait, on each rated trait (henceforth referred to as Δ-ratings). The Δ-ratings were then analyzed in the framework of linear mixed-effects models ("lme4"-package in R [40]), separately for each rated trait. We implemented *a priori* defined planned treatment contrasts comparing the congruent social trait condition (when the trait expression and trait rating coincided, e.g. Δ-ratings of intelligence for intelligence modulations) to ratings of all other social conditions (when the trait expression and the trait rating did not coincide, e.g. Δ-ratings of intelligence for likeability modulations). Thus, we tested directly whether sounds from the congruent trait condition received significantly higher trait ratings on that trait than all other voice modulation conditions. Each model included mean naïve ratings from one of the trait conditions as the outcome variable, the expressed trait as a fixed effect term, and speaker as a random intercept. Likelihood ratio tests were performed to test the effect of trait expression on the Δ - ratings, by comparing the models with fixed effects to the null models with only the random intercepts.

Next, we entered all z-transformed ratings for the neutral voice recordings into a principal component analysis (PCA). This analysis allowed us to explore the underlying dimensions of the social voice space in neutral voices [18]. We then entered all z-transformed Δ-ratings from all modulated voice recordings into the PCA under the hypothesis that vocal modulation would evoke changes in rating behaviour that can be described within the social voice space as a common social navigational map (i.e. exaggerated trait ratings as a function of social vocal modulation). In order to interpret the PCA components, we computed univariate one-way ANOVAs testing the effect of the social vocal control condition on each PCA component. Planned sum contrasts were then computed to test whether the recordings that were obtained from a particular social vocal control condition were significantly different from the overall mean. All principal component analyses were conducted using the package "FactoMineR" in R [41].

Lastly, we aimed to explore the association between basic and social vocal control. Social vocal control performance was operationalized as the mean Δ-rating of a speaker across all social traits. To estimate the basic vocal control performance, we used the average accuracy with which raters classified small or large body-size expressions for each speaker. To test whether the performance in basic vocal control would predict the performance in social vocal control, we computed linear regression models using performance in basic vocal control (mean rate of correctly classified modulated voices across raters for each speaker and averaged over small and large modulation conditions) as the predictor for performance in social vocal control (mean Δ-rating averaged over all congruent trait ratings for each speaker). Due to a recording error, ratings for body-size modulations of one speaker were missing. We also excluded one speaker whose performance in basic vocal control was out of the normal range (with an accuracy below more than 1.5 times interquartile range).

Results

Efficacy in social vocal control

Inter-rater reliability (Cronbach's α) for the modulated voice ratings was .86 for all traits (95% CI: .85 - .87; within trait categories all Cronbach's αs $\geq$ .83, see supplementary table S4). There were significant changes in mean intensity trait ratings evoked by social vocal modulation compared to neutral voices (Δ-ratings) for all expressed traits, all χ^2s(5)> 45.12, all *p*s<.001 (see figure 2A; descriptive results from naïve ratings on neutral voices are reported in supplementary figure S3). To assess the specificity of social vocal control for each trait, we computed planned contrasts to test whether congruent trait Δ-ratings (when the expressed trait and the rated trait coincided) were rated significantly higher than incongruent traits (when expressed traits did not coincide with trait ratings).

Likeable voice modulations evoked a significant mean Δ-rating of .79, *t*(105.90)=4.35, *p*<.001, and were rated as sounding more likeable than any other trait modulation (all *b*s< -.51, all *t*s(200)> -2.9, all *p*s≤ .01). Attractive voice modulations were rated as sounding significantly more attractive than neutral voices (mean Δ-rating =.99, *t*(115.09)=5.89, *p*<.001), and any other trait modulation (all *b*s< -.36, all *t*s(200)> -2.1, all *p*s≤ .042).

Intelligent voice modulations evoked marginally higher ratings in intelligence compared to neutral voices (mean Δ-rating=.28, *t*(95.44)=1.76, *p*=.08) and were perceived as sounding similarly intelligent to confident (*b*=.19, *t*(200)=1.3, *p*=.21), dominant (*b*=-.19, *t*(200)=-1.28, *p*=.20), and likeable voice modulations (*b*=-.12, *t*(200)=-.86, *p*=.39). However, hostile (*b*=-.95, *t*(200)=-6.34, *p*<.001) and attractive (*b*=-.70, *t*(200)=-4.7, *p*<.001) voice modulations, were rated as sounding significantly less intelligent compared to intelligent voice modulations.

Compared to neutral voice samples, confident voice modulations evoked significantly higher ratings on confidence (mean Δ-rating=.67, *t*(100.76)=3.99, *p*<.001). Apart from dominant voice modulations, which did not significantly differ from confident voice modulations on perceived confidence (*b*=-.03, *t*(200)=-.20, *p*=.84), all other trait modulations were rated as significantly less confident (all *b*s< -.33, all *t*s(200)> -2.02, all *p*s≤ .044).

Dominant vocal modulations were perceived as significantly more dominant than neutral voices (mean Δ-rating=1.69, *t*(117.84)=11.21, *p*<.001), and all other trait modulations (all *b*s< -.43, all *t*s(200)> -2.72, all *p*s≤ .007). Lastly, hostile voice modulations evoked a significant mean Δ-rating of 1.93, *t*(152.58)=12.16, *p*<.001, and were perceived as more hostile than all other trait modulations (all *b*s< -1.26, all *t*s(200)> -6.93, all *p*s≤ .001) apart from dominant voice modulations, which were perceived as sounding similarly hostile (*b*=-.28, *t*(200)=-1.53, *p*=.13).

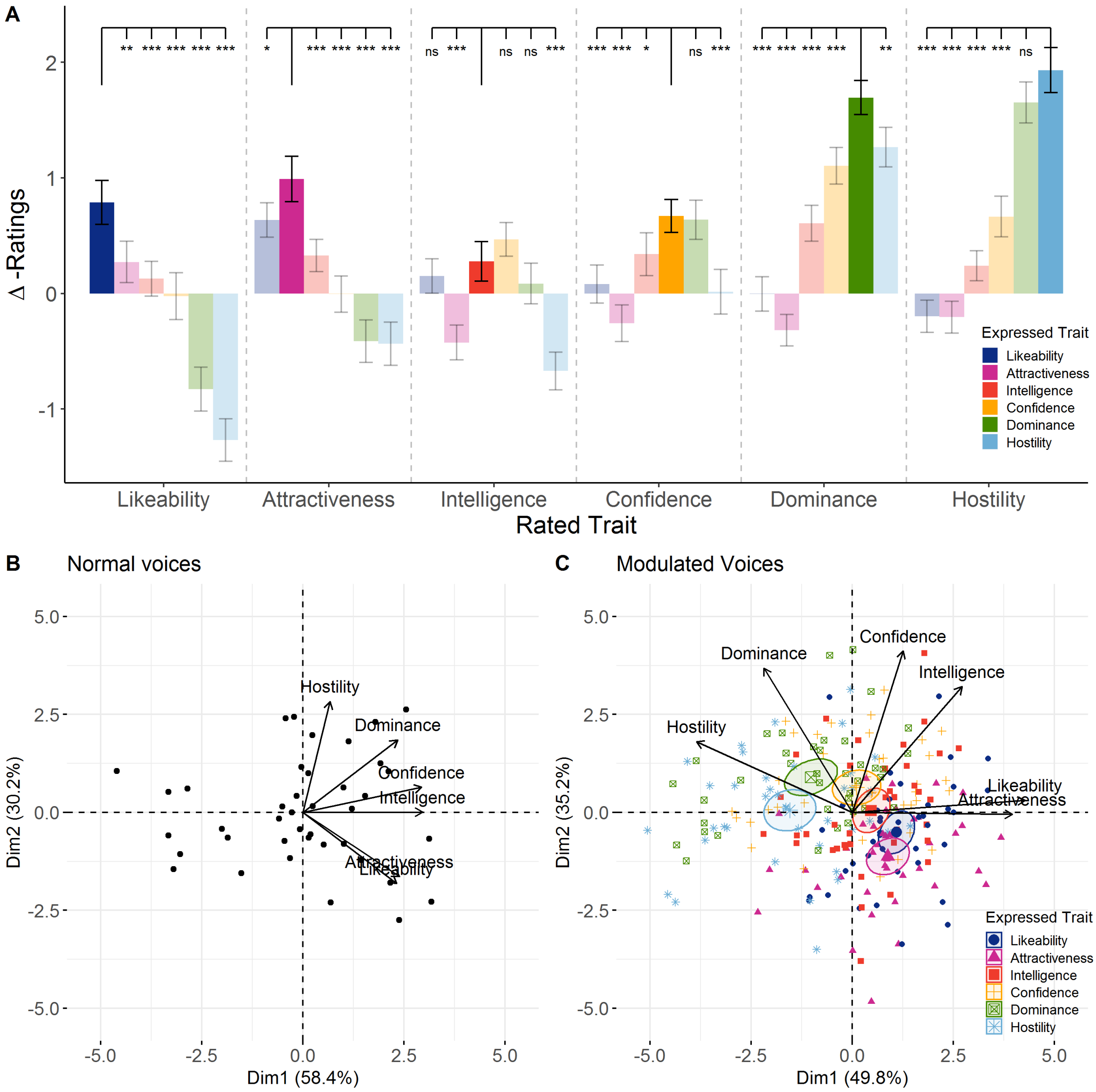


Figure 2. A. Mean changes in trait ratings from normal voices (Δ-ratings) evoked by social vocal control. Planned contrasts show differences in trait ratings over all social vocal control conditions. *** p<.001, ** p<.01, * p<.05. Error bars are standard errors. **B.** Biplot of principal components for normal voice recordings of speakers. **C.** Biplot of principal components for Δ-ratings of modulated voice recordings of those speakers (i.e. ratings of modulated voices above and beyond neutral voices). Ellipses represent 95% confidence intervals around the group means of expressed traits.

Table 1. Principal component loadings of all social traits and explained variance

PCA on …	**ratings of neutral voices**		**Δ-ratings of modulated voices**	
Trait	**PC1**	**PC2**	**PC1**	**PC2**
Likeability	.74	-.58	**.93**	.07
Attractiveness	.76	-.52	**.86**	-.01
Intelligence	**.94**	.00	.59	.70
Confidence	**.94**	.21	.28	**.90**
Dominance	.75	.59	-.48	**.81**
Hostility	.21	**.90**	**-.84**	.40
Explained Variance (%)	58.41	30.25	49.82	35.23

Note. Loadings represent the correlations of the trait judgments with the first two principal components as calculated including all six social traits. Correlations above .8 are highlighted in bold.

Social voice space for modulated voices

Inter-rater reliability for neutral voice recordings was comparable to previous work (Cronbach's α = .85, see supplementary table S4; [18]). We entered all trait ratings obtained for the neutral voice recordings into a principal component analysis to explore the organization of perceived vocal information along the dimensions of the social voice space [18]. The data were adequate for principal component analysis with appropriate inter-correlations (KMO=.73, with individual KMO >.53; Bartlett's test of sphericity, $\chi^2(15)$ = 209.14, p<.001). The first two components cumulatively explained 88.7% of the total variance (PC1 58.41%, PC2 30.25%, PC3 3.1% see Figure 2B and Table 1). Confidence and Intelligence ratings loaded positively and most strongly on the first principal component. For the second component, hostility loaded positively, whereas likeability and attractiveness loaded negatively.

To investigate whether modulated voices represent exaggerations of voices along the social voice space dimensions (affiliation and competence), we next examined the principal components of the changes in trait ratings induced by the modulated voice recordings.

Therefore we entered the Δ-ratings of all recordings of trait modulations into a PCA. The sample contained sufficiently large inter-correlation between items (KMO=.72, with individual KMO >.65; Bartlett's test of sphericity, $\chi^2(15) = 1049.96$, $p<.001$). The first two components cumulatively explained 88.6% of the total variance (PC1 49.8%, PC2 35.2%, PC3 6.1%, see Figure 2C and Table 1). Likeability and attractiveness ratings loaded highly positively on the first component, while hostility ratings showed highly negative loadings. Dominance and confidence ratings loaded most strongly on the second component (see also supplementary materials S6 for PCA of modulated voice based on non-normalized ratings).

The social vocal control condition (i.e. the trait that was expressed) had a significant effect on PC1, $R^2=.31$, $p<.001$. The traits likeability, $b=1.1$, $p<.001$, and attractiveness, $b=0.88$, $p<.001$, were positively associated with PC1. Hostility, $b=-1.54$, $p<.001$, and dominance, $b=-1.02$, $p<.001$, were negatively associated with PC1. For PC2, there was also a significant effect of social vocal control condition on the component with $R^2=.22$, $p<.001$. Here, dominance, $b=0.9$, $p<.001$, and confidence, $b=.63$, $p<.001$, were positively associated with PC2, while likeability, $b=-.51$, $p<.001$, and attractiveness, $b=-1.12$, $p<.001$, were negatively associated.

Efficacy of basic and social vocal control

Individual performance in social vocal control was summarized as each speaker's mean congruent Δ-rating across all six traits (e.g. Δ-rating of intelligence on intelligent voice modulations, Δ-rating of likeability on likeable voice modulation, etc.). Across the speakers, performance ranged from 0.1 to 2.13 with a mean modulation performance – reflected in listener sensitivity to congruent changes - of M=1.06, SD= 0.55 (see supplementary figure S5 A). Thus, for congruent trait modulations, speakers evoked a mean increase of 1 point on the Likert-scale compared to the neutral voice, across all traits.

Basic vocal control was estimated by taking the average accuracy with which large or small voice modulation were correctly classified as such across all raters. Average accuracy over both modulation conditions was 70%. There was a significant difference in accuracy for classification of small voice modulations (M=79%, SD=.16) and large voice modulations (M=61%, SD=.16), $t(37) = -4.46$, $p <.001$ (see supplementary figure S5 B). We found that basic vocal control did not predict performance of social vocal control, adjusted $R^2 = -.03$, $F(1, 36) =.09$, $b = -.31$, $p =.76$.

Discussion

Although aspects of a person's voice are relatively stable across the lifespan [42], the voice can also be highly variable across situational contexts and interactional motives. Although researchers have recently gained more insight into how voices are perceived, the dynamics of volitional vocal modulations in interactions remain largely unexplored. Our study offers novel evidence on the voice as an effective social behaviour. We show that social traits can be successfully *and specifically* expressed through the voice. Moreover, we show that the perceptual changes in social traits evoked by these vocal modulations can be described by two underlying dimensions, namely affiliation and competence. We thus corroborate accounts of the social voice space and extend these findings by showing that speakers can intentionally create amplifications of trait impressions in listeners over and above the trait impressions induced by their neutral voice. Lastly, we show that basic vocal control is not associated with performance in social vocal control, possibly suggesting differential strategies and mechanisms underlying these two types of vocal changes.

Efficacy of social trait expression in voices. Across the expressed social traits, speakers were able to induce an amplification of trait percepts relative to their neutral voice by volitionally modulating their voice while speaking short neutral sentences. These results are in line with previous work showing successful volitional trait expression in voices [15,43]. However, there were differences in the efficacy of voice modulations across social traits: while vocal modulation was particularly successful for expressions of likeability, attractiveness, dominance and hostility, it was less specific for intelligence and confidence voice modulations. In fact, only hostile and attractive voices were clearly discernible from intelligent voices in their perceived intelligence. Hughes et al. [35] report successful expression of intelligence in the voice, reflected in an increase in intelligence ratings when comparing neutral to intelligent voices. Using multivariate ratings, we observed that while intelligent voices were rated significantly higher in intelligence than neutral voices (similar to the findings by Hughes et al. [35]), they are not specific in evoking the desired trait percept (with confident modulations evoking comparable intelligence ratings). This underlines the importance to study not only the sensitivity with which voice changes increase trait perceptions, but also how discernible they are from other voice modulations. This might be relevant for voice modulations in social scenarios where not any voice change leads to a desired outcome, but specific modulations lead to specific trait percepts. Moreover, social interactions involve the sender and the recipient and future work will have to determine the factors that both contribute to the proficiency with which a speaker encodes traits in their voice, and with which the listener decodes these traits accurately.

While intelligence and confidence voice modulations were less specific in evoking respective trait percepts, traits relating to social affiliation were more clearly differentiated. This might be due to the evolutionary significance of this dimension – determining whether an interlocutor has good or bad intentions toward a person will in turn decide whether they should avoid or approach them. The primacy of this dimension for social interactions is reflected in faster response times to affiliation-related words [44], more reliable trait ratings [45], and in competence ratings having only modulatory effect on affiliation ratings [46]. Alternatively, it might be that intelligent and confident voice modulations were more difficult to achieve by speakers [39]; however, we found that speakers were similarly satisfied with all their trait expressions, including intelligence voice modulations (see supplementary figure S2). Lastly, although Mahrholz et al. (2018) have shown high agreement in trait ratings of neutral voices across different types of speech, we cannot rule out that the specific speech content used in our study may have had some influence on the ease with which certain traits could be volitionally expressed and/or perceived (e.g. a sentence about flowers might more easily convey likeability and warmth than intelligence).

In summary, we show that speakers can induce increased trait ratings for a given trait, and that they achieve this specifically for this trait. For instance, expressing likeability leads to increased ratings of likeability, but expressing hostility induces a decrease in ratings of likeability. This might not be surprising, but given that the correlations rely on perceptual differences relative to the neutral voice (Δ-ratings), it speaks to the validity of vocal modulations in achieving trait percepts, particularly in a participant sample that was taken from the general population and not pre-selected to have formal voice training. Importantly, we observed a high inter-rater agreement, similar to previous studies on unmodulated voices [18,24]. Intentional social manipulation of the voice was recognizable and specific, supporting the notion that the voice is a dynamic tool that can be intentionally deployed to express socially relevant information to an interlocutor.

Vocal modulations influence perception in the social voice space. Previous work suggests a social voice space for impression formation [18], similar to that of faces [19]. In our study, ratings of neutral voice samples corroborate this: When judging social traits in neutral voices, listeners ratings could be summarised by two underlying social trait dimensions previously described as the social voice space [18], namely one relating to affiliation (PC2) and one to competence (PC1). The social voice space has been shown to be generalizable over linguistic backgrounds [23] and might emerge through social learning over the course of development [47,48] independent of visual input [49]. Here, we find support for the social voice space for neutral voices for short sentences, in line with previous work noting a high correlation between word and sentence stimulus types when rating social traits [24]. However, in contrast to previous work [18] we also found that the competence dimension explains more variance in the data than the affiliation dimension. The social relevance of our speech exemplars might account for changes in perception particularly of traits relating to the affiliation dimension (e.g. warmth/trustworthiness) [50,51], since we used short ambiguous sentences that were not person-directed or matching with the expressed trait. However, other work has shown that social ratings of voices across ambiguous and socially relevant content is strongly correlated [24], suggesting that the speech content does not significantly influence perceptions of social traits, such as affiliation [52]. Alternatively, our conflicting findings might more likely be an artefact of our instruction in the neutral voice condition to "speak without trying to express anything in particular", thereby potentially toning down information in the voice that is potentially used by listeners to infer how trustworthy or warm a speaker sounds.

We extend previous work by providing novel evidence that trait ratings evoked by intentional voice modulations can also be described by the two primary canonical dimensions of the social voice space. In other words, speakers can specifically amplify the perception of social traits in listeners through intentional vocal modulations and might exploit the social voice space to do so. How vocal modulation interact with listener ratings along the social voice space axes, has previously only been shown for artificial variation of voice pitch, which increased ratings of trustworthiness [53,54] (for comparison, see supplementary table S7 for acoustic profiles of voice modulations in our study), but not for spontaneous human vocalizations. From our findings, we suggest that volitional voice modulation might therefore add to the social make-up used by speakers to situate themselves in the social space, such as in social scenarios where a certain impression is targeted (e.g. to convey competence, confidence, leadership or attractiveness [25–28]). Importantly, these vocal expressions are goal-directed social behaviours, reflecting the attempt to evoke favourable behaviour or judgements from listeners, i.e. to compel a listener to do or think something. Moreover, vocal modulations can be persuasive even if the listener detects that these are deliberate [55]. Additional insights come from populations in whom vocal modulation is changed. For example, patients on the autism spectrum and patients suffering from depression exhibit characteristic acoustic differences during speech along with pervasive social difficulties [56,57]. In these populations, the changed speech pattern can be informative to determine symptom severity and treatment response [58–60].

Modulators of the efficacy of social vocal control. Our data suggest the voice to be a highly dynamic social behaviour that can be strategically modulated to achieve changes in trait impressions. Both modulation tasks evoked specific impressions in listeners: social vocal modulation lead to an average increase of one point on the Likert-scale compared to the neutral voice of a speaker. Also body size modulations in the basic vocal control were effective, particularly for small voice modulations. This is in line with previous work, showing that perceptions of body size can be reliably manipulated through the voice [36]. Interestingly, our results show no significant association between the efficacy of basic and social vocal control. This suggests that vocal control to express social information might be more related by implicit social cognitive processing [39] than explicit vocal control. A speaker might think less about an acoustic target to adjust the voice to, but might rather make use of social scripts with the best outcome prediction. Together with the characteristic prosodic and cognitive empathy deficits in autism spectrum disorders, this suggests that skills related to cognitive empathy, such as perspective taking, contribute to successful social voice modulation. Conversely, high traits of Machiavellianism or psychopathy might positively contribute to vocal modulation success, as shown for facial mimicry [61]. Lastly, both interlocutors - the speakers and the listeners - are important for effective communication [62]. Future studies using interactional set-ups would be useful in understanding dynamic vocal control in more naturalistic social settings [63].

Limitations and future research

In this study, we were interested in the efficacy of expressing social traits in the voice in spite of individual variation in how this might be acoustically achieved [53,54,64,65]. Males and females, for instance, might have different acoustic strategies to signal dominance, with females increasing the harmonics to noise ratio, and decreasing local shimmer and jitter, while there are no observable differences in males in these parameters [35]. Yet, in the same study both males and females are equally effective in evoking increased dominance ratings. Although we did not set out to investigate the acoustic modulators of voice modulation, acoustic analyses might add to the present literature in unveiling which parameters are manipulated by a speaker to encode, and used by the listener to decode social information.

Second, one key aspect of successfully engaging in social vocal control for self-presentation might concern the listener being convinced of a speaker being intelligent, for instance, independently of perceiving that they are trying to sound intelligent. This relates to the authenticity with which speakers are able to express social traits in their voice, which we did not measure in this study. As noted above, variation of nonverbal vocal parameters is effective to persuade others, even if listeners are aware of the persuasion attempt. Moreover, modulated voices even increased the perceived sincerity of the speaker [55]. This might be due to the assumption that nonverbal signals are less controlled and a more honest signal than verbal attempts (for review see [66]). In fact, even if proven to be false, voluntary displays of over-confidence lead to positive person judgements and social status enhancement [67]. To investigate the specific role of nonverbal vocal modulation in self-presentation, future studies should include additional control ratings such as authenticity, valence and arousal to explore the impact of detectable voluntary voice modulation.

Conclusions

We corroborate previous work showing that trait ratings of neutral voices are clustered along the dimensions of the previously reported social voice space. Importantly, we find that exerting social vocal control leads to recognizable and specific changes in the perception of these voices. These changes clustered along the dimensions of affiliation and competence in the social voice space, indicating that speakers can effectively evoke goal-directed trait judgements in listeners over and above their neutral voices.

Data Availability. All de-identified data and supplementary materials can be accessed here: https://osf.io/avkby/?view_only=cd24e889c98448348027f4907a45f09c

Authors' contributions. SG conceived and designed the study, acquired the data, carried out data analysis and wrote the manuscript. CM helped to conceive and design the study, supervised data analysis, and critically reviewed the manuscript. FN helped design the study, supervised data acquisition and critically reviewed the manuscript. HF helped design the study, supervised the work in this study and critically reviewed the manuscript. All authors gave final approval for publication of the manuscript.

Conflict of Interest. We declare we have no competing interests.

Acknowledgements. We thank Alfred Wieland for his assistance in data collection.

Funding. This work was supported by the Forschungsnetz AERIAL (01EE1406C) from the Bundesministerium für Bildung und Forschung to HF. CM is supported by a Research Leadership Award from The Leverhulme Trust (RL-2016-013).

References

1. McGettigan C. 2015 The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication. *Front. Hum. Neurosci.* **9**, 129. (doi:10.3389/fnhum.2015.00129)

2. Krauss RM, Freyberg R, Morsella E. 2002 Inferring speakers' physical attributes from their voices. *J. Exp. Soc. Psychol.* **38**, 618–625. (doi:10.1016/S0022-1031(02)00510-3)

3. Puts DA, Gaulin SJC, Verdolini K. 2006 Dominance and the evolution of sexual dimorphism in human voice pitch. *Evol. Hum. Behav.* **27**, 283–296. (doi:10.1016/j.evolhumbehav.2005.11.003)

4. Sauter DA, Eisner F, Calder AJ, Scott SK. 2010 Perceptual cues in nonverbal vocal expressions of emotion. *Q. J. Exp. Psychol.* **63**, 2251–2272. (doi:10.1080/17470211003721642)

5. Scherer KR. 1972 Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *J. Pers.* **40**, 191–210. (doi:10.1111/j.1467-6494.1972.tb00998.x)

6. Pisanski K, Nowak J, Sorokowski P. 2016 Individual differences in cortisol stress response predict increases in voice pitch during exam stress. *Physiol. Behav.* **163**, 234–238. (doi:10.1016/j.physbeh.2016.05.018)

7. Kreiman J, Sidtis D. 2011 *Foundations of Voice Studies: An Interdisciplinary Approach to Voice Production and Perception*. Wiley-Blackwell. (doi:10.1002/9781444395068)

8. Hellbernd N, Sammler D. 2016 Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *J. Mem. Lang.* **88**, 70–86. (doi:10.1016/j.jml.2016.01.001)

9. Burnham D, Kitamura C, Vollmer-Conna U. 2002 What's New, Pussycat? OnTalking to Babies and Animals. *Science (80-.).* **296**, 1435. (doi:10.1126/science.1069587)

10. Kitamura C, Burnham D. 2003 Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy* **4**, 85–110. (doi:10.1207/S15327078IN0401_5)

11. Dahl A, Sherlock BR, Campos JJ, Theunissen FE. 2014 Mothers' tone of voice depends on the nature of infants' transgressions. *Emotion* **14**, 651–665. (doi:10.1037/a0036608)

12. Pisanski K, Oleszkiewicz A, Plachetka J, Gmiterek M, Reby D. 2018 Voice pitch modulation in human mate choice. *Proc. R. Soc. B Biol. Sci.* **285**, 20181634.

(doi:10.1098/rspb.2018.1634)

13. Anolli L, Ciceri R. 2002 Analysis of the vocal profiles of male seduction: From exhibition to self-disclosure. *J. Gen. Psychol.* **129**, 149–169. (doi:10.1080/00221300209603135)

14. Fraccaro PJ, O'Connor JJM, Re DE, Jones BC, DeBruine LM, Feinberg DR. 2013 Faking it: Deliberately altered voice pitch and vocal attractiveness. *Anim. Behav.* **85**, 127–136. (doi:10.1016/j.anbehav.2012.10.016)

15. Hughes SM, Farley SD, Rhodes BC. 2010 Vocal and physiological changes in response to the physical attractiveness of conversational partners. *J. Nonverbal Behav.* **34**, 155–167. (doi:10.1007/s10919-010-0087-9)

16. Leongómez JD, Binter J, Kubicová L, Stolařová P, Klapilová K, Havlíček J, Roberts SC. 2014 Vocal modulation during courtship increases proceptivity even in naive listeners. *Evol. Hum. Behav.* **35**, 489–496. (doi:10.1016/j.evolhumbehav.2014.06.008)

17. Sorokowski P, Puts D, Johnson J, Żółkiewicz O, Oleszkiewicz A, Sorokowska A, Kowal M, Borkowska B, Pisanski K. 2019 Voice of Authority: Professionals Lower Their Vocal Frequencies When Giving Expert Advice. *J. Nonverbal Behav.* **43**, 257–269. (doi:10.1007/s10919-019-00307-0)

18. McAleer P, Todorov A, Belin P. 2014 How do you say 'hello'? Personality impressions from brief novel voices. *PLoS One* **9**, e90779. (doi:10.1371/journal.pone.0090779)

19. Oosterhof NN, Todorov A. 2008 The functional basis of face evaluation. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 11087–11092. (doi:10.1073/pnas.0805664105)

20. Fiske ST, Cuddy AJC, Glick P. 2007 Universal dimensions of social cognition: warmth and competence. *Trends Cogn. Sci.* **11**, 77–83. (doi:10.1016/j.tics.2006.11.005)

21. Cuddy AJC, Fiske ST, Glick P. 2007 The BIAS Map: Behaviors From Intergroup Affect and Stereotypes. *J. Pers. Soc. Psychol.* **92**, 631–648. (doi:10.1037/0022-3514.92.4.631)

22. Olivola CY, Funk F, Todorov A. 2014 Social attributions from faces bias human choices. *Trends Cogn. Sci.* **18**, 566–570. (doi:10.1016/j.tics.2014.09.007)

23. Baus C, McAleer P, Marcoux K, Belin P, Costa A. 2019 Forming social impressions from voices in native and foreign languages. *Sci. Rep.* **9**, 414. (doi:10.1038/s41598-018-36518-6)

24. Mahrholz G, Belin P, McAleer P. 2018 Judgements of a speaker's personality are correlated across differing content and stimulus type. *PLoS One* **13**. (doi:10.1371/journal.pone.0204991)

25. Schroeder J, Epley N. 2015 The Sound of Intellect: Speech Reveals a Thoughtful Mind, Increasing a Job Candidate's Appeal. *Psychol. Sci.* **26**, 877–891. (doi:10.1177/0956797615572906)

26. Pavela Banai I, Banai B, Bovan K. 2017 Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evol. Hum. Behav.* **38**, 309–314. (doi:10.1016/j.evolhumbehav.2016.10.012)

27. Tigue CC, Borak DJ, O'Connor JJM, Schandl C, Feinberg DR. 2012 Voice pitch influences voting behavior. *Evol. Hum. Behav.* **33**, 210–216. (doi:10.1016/j.evolhumbehav.2011.09.004)

28. Feinberg DR, Jones BC, Little AC, Burt DM, Perrett DI. 2005 Manipulations of fundamental and formant frequencies influence the attractiveness of human male voices. *Anim. Behav.* **69**, 561–568. (doi:10.1016/j.anbehav.2004.06.012)

29. Zuckerman M, Driver RE. 1989 What sounds beautiful is good: The vocal attractiveness stereotype. *J. Nonverbal Behav.* **13**, 67–82. (doi:10.1007/BF00990791)

30. Berry DS. 1990 Vocal attractiveness and vocal babyishness: Effects on stranger, self, and friend impressions. *J. Nonverbal Behav.* **14**, 141–153. (doi:10.1007/BF00996223)

31. Montepare JM, Zebrowitz-McArthur L. 1987 Perceptions of adults with childlike voices in two cultures. *J. Exp. Soc. Psychol.* **23**, 331–349. (doi:10.1016/0022-1031(87)90045-X)

32. McGettigan C, Scott SK. 2014 Voluntary and involuntary processes affect the production of verbal and non-verbal signals by the human voice. *Behav. Brain Sci.* **37**, 564–565. (doi:10.1017/S0140525X13004123)

33. Skoog Waller S, Eriksson M. 2016 Vocal Age Disguise: The Role of Fundamental Frequency and Speech Rate and Its Perceived Effects. *Front. Psychol.* **7**. (doi:10.3389/fpsyg.2016.01814)

34. Morningstar M, Dirks MA, Huang S. 2017 Vocal Cues Underlying Youth and Adult Portrayals of Socio-emotional Expressions. *J. Nonverbal Behav.* **41**, 155–183. (doi:10.1007/s10919-017-0250-7)

35. Hughes SM, Mogilski JK, Harrison MA. 2014 The Perception and Parameters of Intentional Voice Manipulation. *J. Nonverbal Behav.* **38**, 107–127. (doi:10.1007/s10919-013-0163-z)

36. Pisanski K, Mora EC, Pisanski A, Reby D, Sorokowski P, Frackowiak T, Feinberg DR. 2016 Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. *Sci. Rep.* **6**, 34389. (doi:10.1038/srep34389)

37. Kleiner M, Brainard D, Pelli D, Ingling A, Murray R, Broussard C. 2007 What's new in Psychtoolbox-3? *Perception* **36**, 1–16.

38. Brainard DH. 1997 The Psychophysics Toolbox. *Spat. Vis.* **10**, 433–436.

39. Guldner S, Nees F, McGettigan C. 2020 Vocomotor and Social Brain Networks Work Together to Express Social Traits in Voices. *Cereb. Cortex* **00**, 1–17. (doi:10.1093/cercor/bhaa175)

40. Bates D, Mächler M, Bolker BM, Walker SC. 2015 Fitting linear mixed-effects models

using lme4. *J. Stat. Softw.* **67**, 1–48. (doi:10.18637/jss.v067.i01)

41. Lê S, Josse J, Husson F. 2008 FactoMineR: an R package for multivariate analysis. *J. Stat. Softw.* **25**, 1–18.

42. Pisanski K, Fouquet M, Mathevon N, Reby D. 2016 A longitudinal study of within-individual variation in human voice pitch. *J. Acoust. Soc. Am.* **140**, 3397–3397. (doi:10.1121/1.4970892)

43. Tuomi SK, Fisher JE. 1979 Characteristics of simulated sexy voice. *Folia Phoniatr. Logop.* **31**, 242–249. (doi:10.1159/000264171)

44. Ybarra O, Chan E, Park D. 2001 Young and Old Adults' Concerns About Morality and Competence 1. *Motiv. Emot.* **25**.

45. Willis J, Todorov A. 2006 First impressions: Making up your mind after a 100-ms exposure to a face. *Psychol. Sci.* **17**, 592–598. (doi:10.1111/j.1467-9280.2006.01750.x)

46. Wojciszke B, Bazinska R, Jaworski M. 1998 On the dominance of moral categories in impression formation. *Personal. Soc. Psychol. Bull.* **24**, 1251–1263. (doi:10.1177/01461672982412001)

47. Cartei V, Oakhill J, Garnham A, Banerjee R, Reby D. 2020 "This is What a Mechanic Sounds Like": Children's Vocal Control Reveals Implicit Occupational Stereotypes. *Psychol. Sci.* (doi:10.1177/0956797620929297)

48. Cartei V, Cowles W, Banerjee R, Reby D. 2014 Control of voice gender in pre-pubertal children. *Br. J. Dev. Psychol.* **32**, 100–106. (doi:10.1111/bjdp.12027)

49. Oleszkiewicz A, Pisanski K, Lachowicz-Tabaczek K, Sorokowska A. 2017 Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychon. Bull. Rev.* **24**, 856–862. (doi:10.3758/s13423-016-1146-y)

50. Jones BC, Feinberg DR, DeBruine LM, Little AC, Vukovic J. 2010 A domain-specific opposite-sex bias in human preferences for manipulated voice pitch. *Anim. Behav.* **79**, 57–62. (doi:10.1016/j.anbehav.2009.10.003)

51. O'Connor JJM, Barclay P. 2017 The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evol. Hum. Behav.* **38**, 506–512. (doi:10.1016/j.evolhumbehav.2017.03.001)

52. Tsantani MS, Belin P, Paterson HM, Mcaleer P. 2016 Low Vocal Pitch Preference Drives First Impressions Irrespective of Context in Male Voices but Not in Female Voices. *Perception* **45**, 946–963. (doi:10.1177/0301006616643675)

53. Belin P, Boehme B, McAleer P. 2017 The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PLoS One* **12**, e0185651. (doi:10.1371/journal.pone.0185651)

54. Ponsot E, Burred JJ, Belin P, Aucouturier JJ. 2018 Cracking the social code of speech prosody using reverse correlation. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 3972–3977.

(doi:10.1073/pnas.1716090115)

55. Van Zant AB, Berger J. 2020 How the voice persuades. *J. Pers. Soc. Psychol.* **118**, 661–682. (doi:10.1037/pspi0000193)

56. Alpert M, Pouget ER, Silva RR. 2001 Reflections of depression in acoustic measures of the patient's speech. *J. Affect. Disord.* **66**, 59–69. (doi:10.1016/S0165-0327(00)00335-9)

57. Fusaroli R, Lambrechts A, Bang D, Bowler DM, Gaigg SB. 2016 'Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis'. *Autism Res.*, 8 August. , 1–50. (doi:10.1002/aur.1678)

58. Mundt JC, Vogel AP, Feltner DE, Lenderking WR. 2012 Vocal acoustic biomarkers of depression severity and treatment response. *Biol. Psychiatry* **72**, 580–587. (doi:10.1016/j.biopsych.2012.03.015)

59. Yang Y, Fairbairn C, Cohn JF. 2013 Detecting depression severity from vocal prosody. *IEEE Trans. Affect. Comput.* **4**, 142–150. (doi:10.1109/T-AFFC.2012.38)

60. Bone D, Lee C-C, Black MP, Williams ME, Lee S, Levitt P, Narayanan S. 2014 The Psychologist as an Interlocutor in Autism Spectrum Disorder Assessment: Insights From a Study of Spontaneous Prosody. *J. Speech, Lang. Hear. Res.* **57**, 1162–1177. (doi:10.1044/2014_jslhr-s-13-0062)

61. Book A, Methot T, Gauthier N, Hosker-Field A, Forth A, Quinsey V, Molnar D. 2015 The mask of sanity revisited: Psychopathic traits and affective mimicry. *Evol. Psychol. Sci.* **1**, 91–102. (doi:10.1007/s40806-015-0012-x)

62. Hall JA. 1980 Voice Tone and Persuasion. *J. Pers. Soc. Psychol.* **38**, 924–934. (doi:10.1037/0022-3514.38.6.924)

63. Schilbach L. 2016 Towards a second-person neuropsychiatry. *Philos. Trans. R. Soc. B Biol. Sci.* **371**, 20150081. (doi:10.1098/rstb.2015.0081)

64. Ko SJ, Sadler MS, Galinsky AD. 2014 The Sound of Power: Conveying and Detecting Hierarchical Rank Through Voice. *Psychol. Sci.* **26**, 3–14. (doi:10.1177/0956797614553009)

65. Aung T, Puts D. 2020 Voice pitch: a window into the communication of social power. *Curr. Opin. Psychol.* **33**, 154–161. (doi:10.1016/j.copsyc.2019.07.028)

66. DePaulo BM. 1992 Nonverbal Behavior and Self-Presentation. *Psychol. Bull.* **111**, 203–243.

67. Kennedy JA, Anderson C, Moore DA. 2013 When overconfidence is revealed to others: Testing the status-enhancement theory of overconfidence. *Organ. Behav. Hum. Decis. Process.* **122**, 266–279. (doi:10.1016/j.obhdp.2013.08.005)

Supplementary Materials

S1. Vocal Modulation instructions.

Attractive:	Speak as if you were trying to impress someone in whom you are romantically interested.
Dominant:	Speak as if you were trying to assert authority.
Intelligent:	Speak as if they were at a scholarly conference giving a presentation.
Confident:	Speak as if you were trying to make others trust and believe in your ability to do something.
Likeable:	Speak as if you were trying to be liked by someone else.
Hostile:	Speak as if you want to distance yourself from someone else.
Large:	Speak as if you were physically taller than you are.
Small:	Speak as if you were physically smaller than you are.

Figure S2. Speakers' mean performance ratings of voice modulations A. For all recordings. B. for recordings selected for naïve ratings. Speakers own performance ratings were obtained after each trail and were given on a visual analogue scale with anchors at 0 (=not at all) and 100 (=very). For each expressed trait and speaker, only those recordings that received maximal intensity ratings was passed on to naïve listeners for ratings. **A.** $Mean_{Attractiveness}$=65.1, SD=23.2; $Mean_{Dominance}$=69.4, SD=22.1; $Mean_{Hostility}$=72.2, SD=21.4; $Mean_{Intelligence}$=68.3, SD=21.0; $Mean_{Confidence}$=67.4, SD=20.7; $Mean_{Likeability}$=71.6, SD=20.0; $Mean_{Neutral}$=86.8, SD=15.2. **B.** $Mean_{Attractiveness}$=82.1, SD=16.1; $Mean_{Dominance}$=85.8, SD=15.6; $Mean_{Hostility}$=87.8, SD=13.8; $Mean_{Intelligence}$=85.9, SD=14.3; $Mean_{Confidence}$=84.9, SD=15.2; $Mean_{Likeability}$=87.1, SD=14.1; $Mean_{Neutral}$=95.5, SD=8.96.

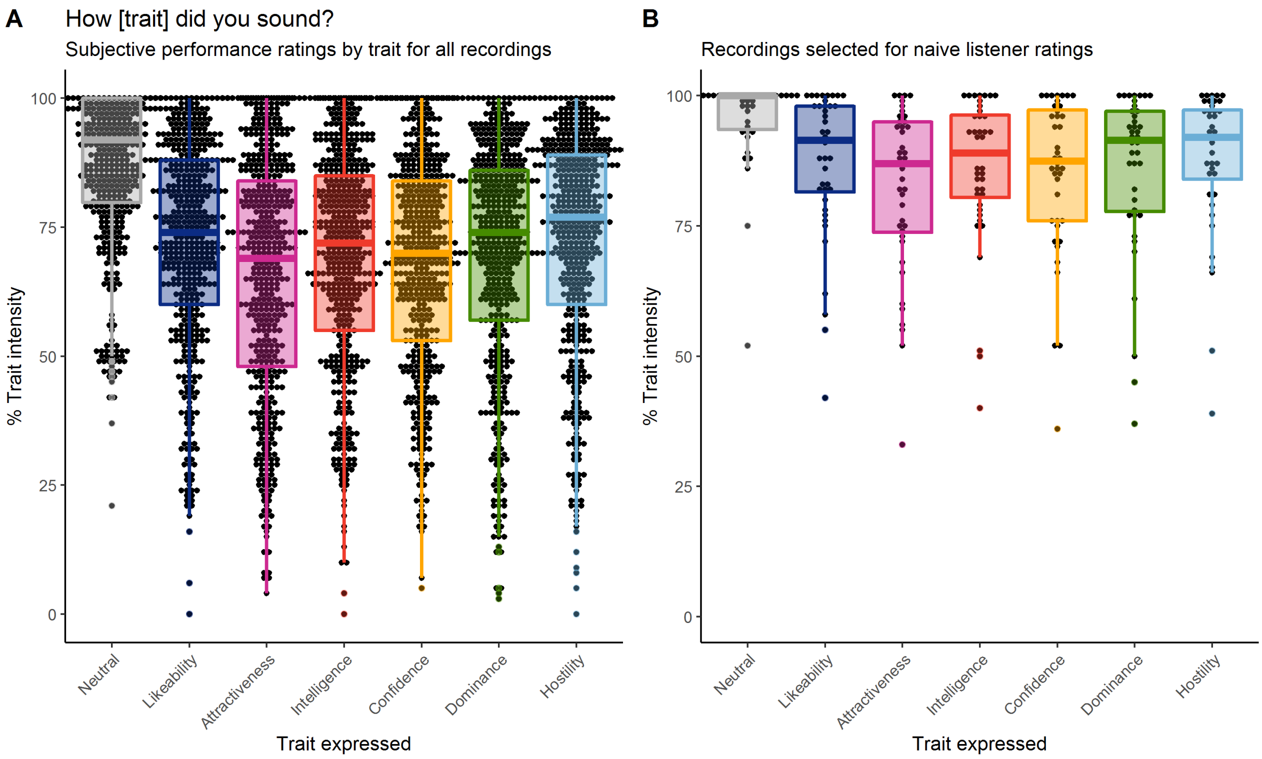
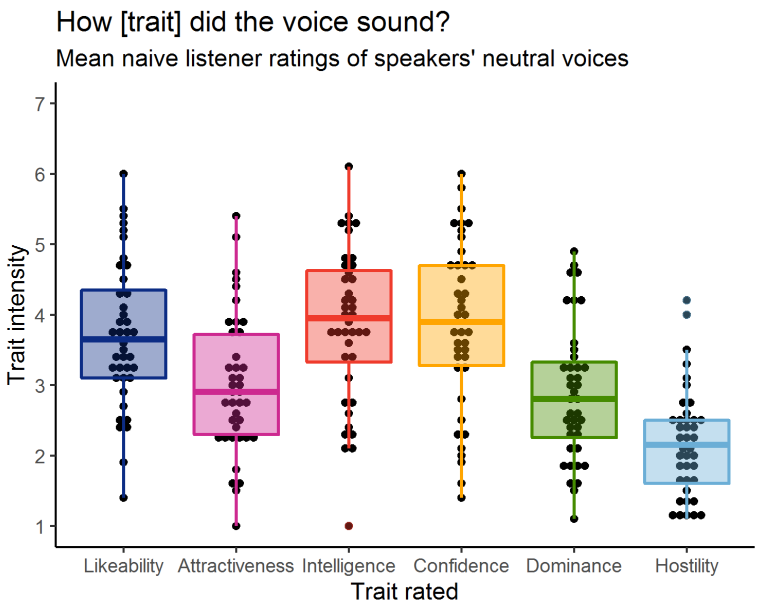


Figure S3. Mean ratings of neutral voices. Naïve listener ratings obtained for the neutral voice recordings on all traits for each speaker given on a 7-point Likert-scale ranging from 1 (=not at all) to 7 (very). $Mean_{Attractiveness}$=3.01, SD=0.99; $Mean_{Dominance}$=2.88, SD=0.96; $Mean_{Hostility}$=2.19, SD=0.76; $Mean_{Intelligence}$=3.85, SD=1.10; $Mean_{Confidence}$=3.86, SD=1.20; $Mean_{Likeability}$=3.73, SD=1.03.

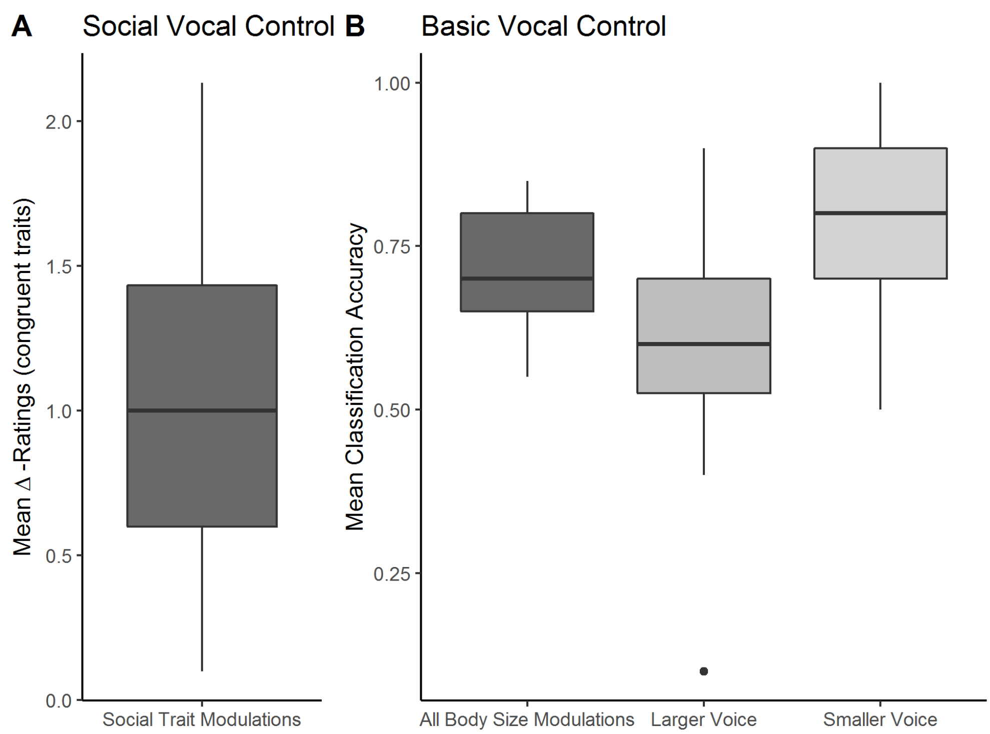


S4. Inter-rater reliability. Cronbach's α for normal voice recordings and expressed traits.

Table S4. Cronbach's α scores grouped by each expressed trait and the normal voice ratings among the 10 raters of each recording.

Trait	**Cronbach's α**	**95% CI (lower - upper)**
Likeability	.87	.84 - .89
Attractiveness	.86	.84 - .89
Intelligence	.83	.8 - .87
Confidence	.85	.82 - .88
Dominance	.86	.84 - .89
Hostility	.86	.84 - .89
Normal	.85	.83 - .85
All recordings	.86	.85 - .87

Figure S5. Vocal Control Performance. A. Social Vocal Control. Mean Δ-ratings across all congruent trait expressions and ratings. A higher value signifies a larger deviation from the normal voice rating (an increase in intelligence ratings for intelligent voice modulations relative to the intelligence rating for the neutral voice) **B. Basic Vocal Control.** Average accuracy over all raters for vocal body size modulations as the relative number of correctly classified recordings shown for each speaker. A value of 0.5 is equal to 50% of raters classified the modulated recording correctly.



S6. Results of PCA of mean ratings on social traits for modulated voices. The sample contained significantly large enough inter correlation between items to allow PCA analysis (KMO=.72, with individual KMO >.63; Bartlett's test of sphericity, $\chi 2(15)$ = 1345.10, p<.001). The first two components cumulatively explained 87% of the total variance (see S5 Table 1). **Figure S6** depicts a biplot of loadings on mean ratings of social traits on modulated voices grouped by the social traits expressing the voice including group centroids. Ellipses illustrate 95% confidence intervals on group centroids of social traits expressed in the voices. **Table S6** shows the loadings of social trait ratings on the two first principal components.

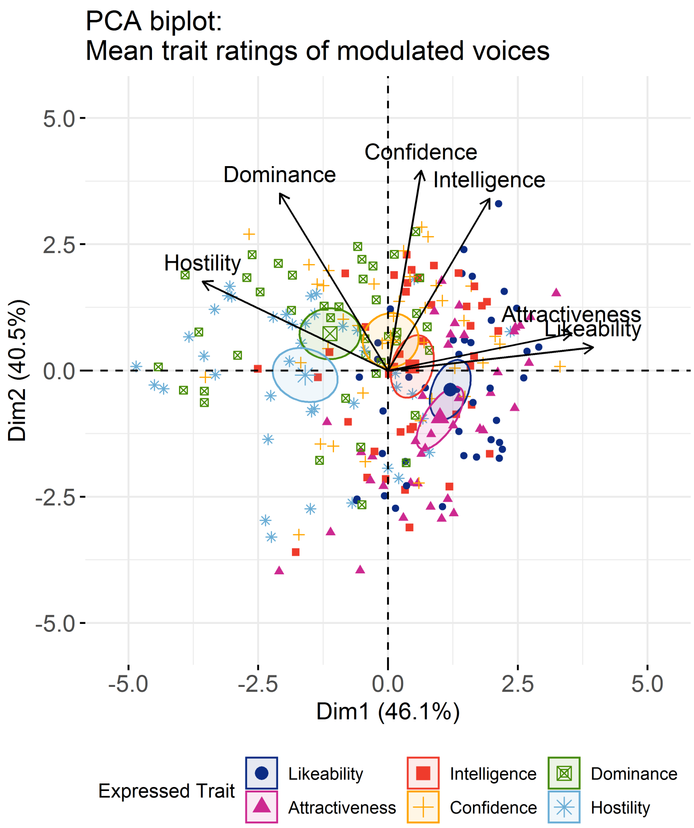


Table S6. Principal component loadings of all social traits and explained variance on modulated voice ratings

Trait	PC1	PC2
Likeability	**.9**	.3
Attractiveness	.81	.28
Intelligence	.25	**.91**
Confidence	.17	**.95**
Dominance	-.66	.71
Hostility	**-.92**	.21
Explained Variance (%)	46.1	40.5

Note. Loadings represent the correlations of the trait judgments with the first two principal components as calculated including all six social traits. Correlations above .9 are highlighted in bold.

Table S7. Acoustic description of normal and modulated voices.

Expressed Trait	**Duration (s)**	**Mean F0 (Hz)**	**SD F0 (Hz)**	**Mean Intensity (dB)**	**% Unvoiced**	**HNR**	**Spectral centre of gravity (Hz)**	**SD spectrum (Hz)**
Normal	1.75 ± .32	205.96 ± 52.45	84.71 ± 43.28	77.82 ± 3.28	28.59 ± 11.41	14.20 ± 3.33	527.99 ± 290.46	1178.46 ± 766.4
Likeable	1.65 ± .29	213.61 ± 50.71	90.54 ± 30.37	78.66 ± 2.99	27.53 ± 10.8	14.38 ± 3.33	521.99 ± 260.54	1176.83 ± 743.6
Attractive	1.82 ± .28	194.41 ± 48.5	81.00 ± 33.32	76.85 ± 3.65	26.12 ± 10.93	14.47 ± 3.02	416.52 ± 242.89	934.14 ± 604.75
Intelligent	1.91 ± .39	208.40 ± 51.29	80.52 ± 37.25	79.61 ± 3.19	28.78 ± 12.53	15.11 ± 3.59	528.96 ± 362.28	1168.91 ± 807.44
Confident	1.68 ± .26	206.67 ± 50.3	78.71 ± 36.73	80.50 ± 3.48	28.35 ± 11.12	14.47 ± 3.61	571.18 ± 258.71	1294.10 ± 691.39
Dominant	1.80 ± .31	195.64 ± 43.25	68.06 ± 31.53	81.56 ± 3.41	29.42 ± 11.43	14.41 ± 3.45	592.32 ± 320.51	1316.36 ± 766.57
Hostile	1.74 ± .32	192.13 ± 38.75	75.72 ± 42.98	79.86 ± 4.63	31.24 ± 10.91	13.39 ± 3.23	594.37 ± 317.13	1445.72 ± 709.73
Large	2.08 ± .36	192.53 ± 51.6	70.09 ± 32.58	80.67 ± 4.14	27.97 ± 10.83	15.25 ± 3.61	508.46 ± 283.42	1060.69 ± 636.82
Small	1.95 ± .36	259.89 ± 95.63	86.36 ± 38.40	75.81 ± 3.8	32.89 ± 11.18	14.93 ± 3.7	504.28 ± 234.7	1138.36 ± 623.22

Note. SD = standard deviation, HNR = Harmonics-to-noise-ratio. Extraction of acoustic parameters was done using the software package PRAAT (Boersma & Weenink, 2015).

References for Supplementary Materials

Boersma, P., & Weenink, D. (2015). Praat: doing phonetics by computer. Version 6.0. 15.

2.2 Study 2: Vocomotor and social brain networks work together to express social traits in voices[2]

[2] Guldner, S., Nees, F., & McGettigan, C. (2020). Vocomotor and Social Brain Networks Work Together to Express Social Traits in Voices. *Cerebral Cortex*, *30*(11), 6004–6020. https://doi.org/10.1093/cercor/bhaa175

Vocomotor and social brain networks work together to express social traits in voices

Stella Guldner[1,2,3*], Frauke Nees[1,4] and Carolyn McGettigan[3,5*]

[1] Department of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2] Graduate School of Economic and Social Sciences, University of Mannheim, Mannheim, Germany

[3] Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK

[4] Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany

[5] Department of Psychology, Royal Holloway, University of London, Egham, UK

Corresponding authors:

Carolyn McGettigan

Department of Speech, Hearing and Phonetic Sciences

University College London

2 Wakefield Street

London WC1N 1PF, UK

Telephone: +44 (0)20 7679 4075

Email: c.mcgettigan@ucl.ac.uk

Stella Guldner

Institute of Cognitive and Clinical Neuroscience

Central Institute of Mental Health

J5, 68159 Mannheim, Germany

Telephone: +49 (0) 621 1703 6314

Email: stella.guldner@zi-mannheim.de

Abstract

Voice modulation is important when navigating social interactions – tone of voice in a business negotiation is very different from that used to comfort an upset child. While voluntary vocal behaviour relies on a cortical vocomotor network, social voice modulation may require additional social cognitive processing. Using functional MRI, we investigated the neural basis for social vocal control and whether it involves an interplay of vocal control and social processing networks. Twenty-four healthy adult participants modulated their voice to express social traits along the dimensions of the social trait space (affiliation and competence), or to express body-size (control for vocal flexibility). Naïve listener ratings showed that vocal modulations were effective in evoking social trait ratings along the two primary dimensions of the social trait space. Whereas basic vocal modulation engaged the vocomotor network, social voice modulation specifically engaged social processing regions including the medial prefrontal cortex, superior temporal sulcus and precuneus. Moreover, these regions showed task-relevant modulations in functional connectivity to the left inferior frontal gyrus, a core vocomotor control network area. These findings highlight the impact of the integration of vocal motor control and social information processing for socially meaningful voice modulation.

Keywords: fMRI, social communication, social traits, vocal control, voice production

Vocomotor and social brain networks work together to express social traits in voices

Similar to the impression management we perform when we dress up for a job interview, can we influence how we are perceived by volitionally modulating our voice? Despite the importance of the voice for social judgement formation and the voice being a fundamentally social behaviour (McGettigan 2015), we understand little about the cognitive and neurobiological underpinnings of expressing social information in voices. Although the voice remains relatively stable within its anatomical boundaries throughout adult life (Pisanski, Fouquet, et al. 2016), vocal behaviour is adaptable to specific social contexts and expressing socially relevant information (reviewed by Pisanski et al., 2016). Consequently, the voice carries a multitude of information about a speaker, such as their emotional state and personality traits (Banse and Scherer 1996; Krauss et al. 2002; Sauter et al. 2010; Pisanski, Cartei, et al. 2016; Oleszkiewicz et al. 2017). Not surprisingly, vocal information is spontaneously used by listeners to infer the intentions of an interlocutor (Hellbernd and Sammler 2016), e.g. to judge how trustworthy or dominant a speaker is (McAleer et al. 2014). These judgements can be represented in a social voice space (McAleer et al. 2014), showing spontaneous attribution of traits signalling approach or avoidance, i.e. how likeable or dislikeable a speaker might be, and how socially potent they are, i.e. how intelligent they sound. The social voice space is stable across cultures (Baus et al. 2019) and related to specific acoustic modulation patterns. Pitch contour, for instance, is closely associated with ratings of trustworthiness (Belin et al. 2017; Ponsot et al. 2018), while an interaction of both pitch and intensity measures has been related to expressing hierarchy or confidence in the voice (Ko et al. 2014; Jiang and Pell 2017).

The two dimensions of the social voice space, namely affiliation (also termed warmth or trustworthiness) and competence (also termed dominance or confidence) represent the axes of the social trait space (Fiske, Cuddy, & Glick, 2007; Harris & Fiske, 2007), parallel to the previously reported social space of faces (Todorov et al. 2005; Todorov, Said, et al. 2008). Receiving beneficial judgements in the social space is important to achieve successful interactions: previous work suggests that vocal cues might be an important contributor to this, for example predicting positive outcomes in a job interview (Schroeder and Epley 2015) or a political election (Pavela Banai et al. 2017). Dynamic voice changes can be observed spontaneously in response to external cues (e.g. raising the volume of one's voice in a noisy environment), as well as intentionally in response to internal goals (e.g. trying to impress a panel at an interview). Related to the latter context, vocal control describes the capacity to perform goal-directed and voluntary modulation of suprasegmental speech characteristics during voice production. Through vocal control, speakers can influence how old (Skoog Waller & Eriksson, 2016), or how feminine versus masculine they are perceived (Cartei, Cowles, Banerjee, & Reby, 2014). Moreover, immediate social information can also be communicated through controlled vocal modulations, such as social emotions (Morningstar et al. 2017). Such voluntary voice modulation to express social information, i.e. *social vocal control*, can directly impact listeners, who make use of vocal information to make spontaneous social trait judgments. Hughes and colleagues (2014) recorded speakers' voice modulations to express traits, such as dominance, and presented them to naïve listeners. Compared to the speakers' normal (i.e. non-modulated) voice recordings, dominant voices, for instance, were indeed rated as higher in dominance than normal voices. This study suggests that social vocal control presents an effective interpersonal tool, which can be instrumental in eliciting beneficial social judgements. However, Hughes and colleagues (2014) only ever obtained perceptual ratings on the *intended* trait for each modulated voice category (e.g. speech intended to sound dominant was only rated for dominance and not, for example, for trustworthiness). Hence, the specificity, and therefore the potency, of social vocal modulations remains unclear.

Exerting vocal control has been shown to rely on a fronto-parietal vocomotor control network (VMN) between the IFG pars triangularis/opercularis, supplementary motor area (SMA), the supramarginal gyrus (SMG), insula, superior temporal cortex (STC), anterior cingulate cortex (ACC), basal ganglia (BG) and cerebellum (reviewed by Pisanski, Cartei, et al., 2016; Simonyan & Horwitz, 2011). The IFG has a crucial role in speech motor control during vocalization. It is thought to be a central executive and primary input region for voluntary voice production (Hage and Nieder 2016), representing speech sound maps for feed-forward vocal control, particularly in opercular parts of left IFG (Tourville and Guenther 2011). Thus, it provides input to primary motor cortex, which in turn engages a cortical and subcortical network to exert control over vocal production (Simonyan and Horwitz 2011).

To our knowledge, no study has specifically targeted the neurobehavioural mechanisms involved in *social* vocal expression. Some insight comes from studies investigating vocal modulation to express affect. The expression of such affective vocalizations has been proposed to rely on the interaction of a dual-pathway system consisting of the neocortical regions of the VMN and a phylogenetically older network of subcortical brain structures such as the basal ganglia and the amygdala (Ackermann et al. 2014; Hage and Nieder 2016). In line with this, voluntary affective vocal expression engages both vocomotor areas related to volitional expression as well as areas related to processing affect, such as the IFG, BG, ACC and STC and Amygdala (Barrett et al. 2004; Aziz-Zadeh et al. 2010; Laukka et al. 2011; Pichon and Kell 2013; Frühholz et al. 2015; Klaas et al. 2015; Belyk and Brown 2016; Mitchell et al. 2016; Klasen et al. 2018). This interplay of affect processing streams and the vocomotor network therefore suggests that some informational integration, is necessary to achieve the successful expression of affect in the voice.

In line with this, studies investigating the association between social traits and voice production point to an interaction of social processing areas with vocal motor processing areas during socially meaningful voice changes. Klasen and colleagues (2018) showed that activation in the right superior temporal cortex (STC), was modulated by the social context during emotional voice production, suggesting that areas associated with domain general social processing (social brain network; SBN) might be involved in expressing socially relevant information in voices during speech, as it is for perception of social traits in voices (Hellbernd and Sammler 2018). Another study of speech production that directly manipulated third person vocal identity expression (i.e. impersonations) reported activated regions in the right superior temporal sulcus (STS) that were functionally connected to left IFG during voice production (McGettigan et al. 2013). In another study, actors gave improvised, but covert, answers to questions about themselves while in role. In line with the previous studies, responding in the first person of a fictional character (during acting) also engaged right posterior STS regions in addition to a network of ventral (vmPFC) and dorsal medial prefrontal regions (dmPFC) and precuneus, which are all implicated in the SBN (Brown, Cockett, & Yuan, 2019). Although less specifically targeting vocal modulation in their task, the latter study is one of the few that offers some insight into intentional first person trait modulation, and supports the idea that SBN regions might also be instrumental in achieving voluntary expressions of identity in the voice.

Social vocal control is an intentional goal-directed behaviour that requires a socially beneficial expression of self-related traits. To be successful, social vocal control should therefore involve some form of trait processing. Together with the medial prefrontal cortex (mPFC), the posterior portion of the STS (or temporo-parietal junction; TPJ) is an important and domain-general contributor to the social brain network (Van Overwalle 2009; Bzdok et al. 2012; Schurz et al. 2014) and is engaged during evaluative judgments of affective information in voices (Dricu and Frühholz 2016). The mPFC is reliably activated during tasks that require mentalizing (Schurz et al. 2014) and has been proposed to subserve social trait judgment along the dimension of the social trait space (Harris et al. 2005; Harris and Fiske 2007; Ma et al. 2013, 2016; Van Overwalle et al. 2016), particularly in ventral parts (Harris and Fiske 2007; Van Duynslaeger et al. 2007; Ma et al. 2014, 2016; Tavares et al. 2015; Van Overwalle et al. 2016). Moreover, mPFC is specifically engaged during tasks requiring psychological self-representation, possibly reflecting emotional or evaluative processing of the conceptual self. Compared to other-referential tasks, self-referential tasks are reliably associated with increased activation in SBN areas, such as mPFC, bilateral STG, precuneus, and TPJ (Hu et al. 2016). In fact, mPFC and pSTS/TPJ are involved in trait processing both when social traits are in reference to oneself or to another person (Nicolle et al. 2012). While the mPFC represents rather long-term, static trait information about others (and likely, the self), the pSTS/temporo-parietal junction is involved in rapid, short-term intention and goal attribution (Saxe and Powell 2006; Ma et al. 2012) both for verbal and nonverbal information (Redcay 2008; Shultz et al. 2012; Redcay et al. 2016). In summary, social vocal control to express beneficial traits might entail an integration of vocomotor and social trait processing areas in relation to the self. Although social trait perception in voices has recently gained attention (e.g. Hellbernd & Sammler, 2018), no study to our knowledge has tested this directly in voice production.

Lastly, although some evidence points to the representation of a common trait code in the ventral mPFC (Van Overwalle et al. 2016), the question of whether the two main dimensions of the social trait space engage separable neural regions remains controversial. While the affiliative traits (e.g. warmth, trustworthiness) are associated with processing in ventral mPFC, competence evaluations have been shown to additionally engage the precuneus (Ma et al. 2016). In the present study, both competence and affiliative traits were expressed volitionally in the voice in our social vocal control task, allowing insights on the neural representation of these two social space dimensions.

In summary, although speaking can be understood as a goal-directed social behaviour, little is known about how social traits are encoded in the voice. The current study addressed this gap in the literature, posing two research questions on the issue. First, we tested whether social vocal control would be effective in evoking percepts that vary along the dimension of the social trait space: affiliation and competence. Here, we tested in particular beneficial social traits, i.e. sounding likeable or hostile, and sounding intelligent to navigate interactions. This is in alignment with the spontaneous perception of voices on the social voice space (McAleer et al. 2014) and comparable to commonly expressed social information in everyday social interactions. Second, we aimed to illuminate the functional neural correlates supporting social vocal control to navigate this social space, i.e. to express social traits. Based on the literature on neural correlates of basic vocal control, we predicted significant changes in neural activation during vocal modulation in the VMN (including the pars triangularis and pars opercularis of the left IFG as a central vocal control area, as well as supplementary motor area, basal ganglia, cerebellum, and insula). For social vocal control, we expected a higher engagement of areas involved in modality-independent social trait and self-referential trait processing, such as pSTS/TPJ and mPFC. To achieve social vocal control, we expected that the core vocal control network (left IFG) would be functionally connected to social processing areas during online social voice modulation. According to accounts of the social trait space, we further explored whether social vocal control to express traits along the two dimensions of affiliation and competence would rely on differential engagement of social processing areas and lastly, whether there would be differential functional connectivity profiles with left IFG for these two dimensions of the social trait space.

Methods

Participants. Twenty-four right-handed, native British English speakers (M_{age} = 21.04 SD=3.26, 3 male) participated in this experiment. All participants had normal or corrected-to-normal vision and reported no history of hearing, language, neurological or psychiatric disorders. Volunteers were recruited from the participant pool at the Department of Psychology at Royal Holloway, University of London and received 30 GBP as reimbursement. All participants provided their full informed and written consent prior to participation according to the Declaration of Helsinki (1991; p. 1194). This study was approved by the research ethics committee of Royal Holloway, University of London (587-2017-10-24-14-50-UXJT010).

Social vocal control task. The main experimental task consisted of a social vocal control task in which participants were asked to express social and non-social traits in the voice. The social traits were selected to represent the two principal dimensions of the social trait space (Fiske et al. 2007), namely affiliation (also warmth/trustworthiness) and competence (also referred to as dominance; see Belin, Boehme, & McAleer, 2017; McAleer, Todorov, & Belin, 2014 for prior work on trait dimensions in voices). Social traits included vocally expressed intelligence (competence), and likeability and hostility (affiliation dimension), thus spanning the social voice space reflecting spontaneous social appraisal of neutral voices (McAleer et al. 2014). Modulating the voice to express a large body size, as well as speaking in non-modulated "normal voice", were implemented as control conditions (see supplementary materials S1). A body-size modulation was chosen as a control condition because it demands substantial vocal tract manipulations (e.g. larynx lowering, changes in vocal fold tension), without having direct social trait implications (Pisanski, Mora, et al., 2016). Exemplars consisted of four two-syllable, five-letter pseudowords with a C-V-C-V-C (C=consonant, V= vowel) phonotactic structure (i.e. *belam*, *lagod*, *minad*, and *namil*; Frühholz et al. 2015).

These exemplars have previously been used in tasks involving voice production with affective content (Frühholz et al., 2015).

Design and procedure. Prior to the main experiment, participants completed a short training task, during which they were introduced to the exemplars and social traits to be expressed and could familiarize themselves with the social vocal control task. In the scanner, participants then completed the task, which consisted of 4 runs of 150 trials, including 30 rest trials. Each run consisted of a randomized order of the 5 vocal modulation conditions (normal / large / hostile / likeable / intelligent) paired with one of the 4 exemplars. Each exemplar and condition combination was repeated 6 times over the course of a run. Out of these, 3 repetitions were Go trials and 3 were No-Go trials. Go and No-Go trials were presented in randomized order, with the restriction of a maximum of three No-Go trials in a sequence. Within each trial, participants were presented with a social target trait and a fixation cross for two seconds. During this time they were asked to prepare to express the target trait in their voice. In Go trials, the exemplar was then shown at the position of the fixation cross for 1.5 seconds. At the beginning of this silent gap the exemplar appeared on the screen and participants were asked to vocalize the exemplar while expressing the target trait. During No-Go trials the fixation cross remained on the screen for the duration of the silent gap and no exemplar was presented (see Figure 1). We included both Go and No-Go trials in the task, to filter neural activation specifically related to active, ongoing voice production (Go) and thereby exclude any other task-related cognitive processing effects (present in Go and No-Go). Importantly, participants prepared to speak in both trials, but only received a target word in Go trials.

Vocal recordings of speech in Go trials were recorded using an in-scanner MR-compatible microphone (Opto-acoustics, FOMRI-III). Visual cues were presented using the

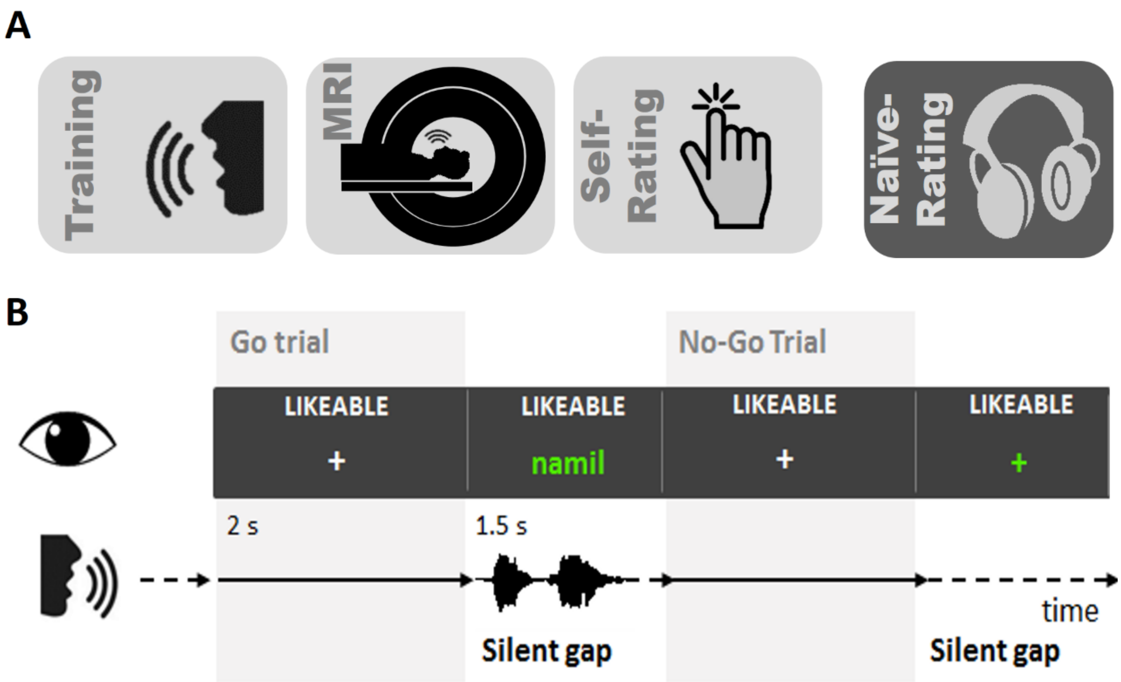


Figure 2. Experimental procedure and example trial structure of the social vocal control task. **A**. Participants first completed a training session, prior to completing the social vocal control task in the scanner. Post-scanning, each speaker rated all of their own vocal recordings on 7-point Likert-scales (self-rating). Lastly, naïve listeners rated the social vocal control performance of each speaker's best (i.e. self-selected) vocal modulations on all social traits. **B**. During the scanning sessions, participants prepared each social trait expression for 2 s before being shown the relevant exemplar. During Go trials, speech was recorded during the silent inter-scan gap of 1.5 s after the exemplar was presented. During No-Go trials, no exemplar was presented, and the fixation cross remained on the screen but changed its color to green to indicate the duration of the silent gap.

Psychophysics toolbox (Brainard, 1997, Kleiner, Brainard & Pelli, 2007) in Matlab (2014a, the Mathworks, Natick, MA) – these were projected onto a screen at the back of the scanner bore and viewed via a mirror on the headcoil. The total scanning time was around 50 minutes.

After scanning, participants were re-invited to rate their own performance on the social vocal control task, using the voice recordings acquired during the scanning session. To do this, they were presented with their voice recordings blocked by the expressed target trait. Recordings were presented in a soundproof booth via Sennheiser Headphones HD306, using the Psychophysics toolbox (Brainard 1997, Kleiner, Brainard & Pelli, 2007) in Matlab (2014a, the Mathwork, Natick, MA). For each recording the participant was re-invited approximately 1 week later and asked to evaluate how strongly they had expressed the social trait on a 7-point Likert scale ranging from not at all (=1) to very much (=7). Based on these subjective ratings, we selected one recording for each trait category, for each participant that had received the participant's own maximal rating for that trait. This was done to ensure that independent evaluations of performance (i.e. with naïve listeners) would be based on sounds where the original participants had felt confident in their accurate expression of the target trait. In cases where multiple recordings had the same maximal rating, we selected one token at random.

Naïve ratings. For each participant the highest rated recording for each trait category was then presented to naïve listeners to obtain a performance index. For this task, the normal voice recordings were intensity normalized across all 24 speakers. The resulting normalization parameter was then used to normalize trait recordings within each speaker: this preserved within-speaker differences between the traits while maintaining average intensity levels between speakers. The normalized recordings were then presented to 24 naïve listeners (M_{age}=19.92, SD=1.47, 4 male). Listeners were recruited through at the Department of Psychology at Royal Holloway, University of London. All raters signed their informed consent prior to participation and received 5 GBP as reimbursement for their participation.

To reduce experiment duration and avoid fatigue effects, each rater heard only a subset of 10 speakers. The subset of speakers was counterbalanced to ensure that each speaker was heard by at least 10 different listeners. Recordings were presented in randomized order within 7 blocks. Each block consisted of one of the 4 relevant social trait modulations to be rated (hostility, likeability, intelligence and body size), or an additional stimulus property (i.e. arousal, valence, authenticity). One recording of each speakers' normal voice was included in the task, since we were interested how the trait ratings induced by the vocal change differed from the normal voice of each speaker. Using this approach, each listener rated each recording on all traits over the course of the blocks: Ratings were given on 7-point Likert scales measuring how strongly the heard voices expressed a given trait, ranging from not at all (=1) to very (=7). For the valence ratings, the scale ranged from very negative (=-3) to very positive (=3). Arousal ratings were rated on a Likert-scale ranging from very sleepy (=1) to very alert (=7). Listeners also rated each recording on perceived authenticity – however, due to some uncertainty about how participants interpreted this scale we have chosen not to include the authenticity ratings in the analyses reported here. The order of blocks was randomized between listeners. Stimuli were presented using the Psychophysics Toolbox (Brainard 1997, Kleiner, Brainard & Pelli, 2007) in Matlab (2014a, the Mathwork, Natick, MA) on Sennheiser Headphones (Sennheiser U.K. Ltd, Marlow, UK) in a soundproof booth.

fMRI acquisition and analysis

Task-based fMRI. Functional brain images were acquired on a 3T Siemens TIM Trio scanner with a 32 channel headcoil, using a rapid–sparse event-related 3D echo-planar imaging (EPI) sequence (32 axial slices, slice gap 25%, resolution 3x3x3mm^2, flip angle 78°, matrix 64 x 64, TE: 30 msec, TR: 3.5 sec, TA: 2 sec). A 3D T1-weighted MP-RAGE scan was acquired for EPI image alignment and spatial normalisation (voxel size 1 mm isotropic; flip angle 11°; TE 3.03 ms; TR 1830 ms; image matrix 256 x 256). Analysis was conducted in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). Preprocessing steps included spatial realignment, segmentation, co-registration, normalization (functional images were resampled to a voxel size of 2x2x2mm) and smoothing (FWHM=8mm). 1st Level general linear models included the conditions as regressors and subjective ratings as parametric modulators, which were analysed in the framework of one-sample T-tests at the second level. We used a significance threshold of $p<.001$ for second level tests, uncorrected for multiple comparisons. To ensure a type 1 error of $p=.001$ at the individual voxel level and a threshold of $p=.05$ corrected for multiple comparisons at the cluster level, a cluster extent threshold was computed for each contrast. This threshold was determined using 1,000 Monte-Carlo simulations based on whole-brain fMRI activation, as described elsewhere (Slotnick et al. 2003). The resulting clusters were labelled based on the location of each peak activation using the in-built Neuromorphometrics and the automated anatomical labeling (AAL) atlas in SPM12. For illustrations, parameter estimates were extracted from significant clusters in the group maps using the MarsBar Toolbox in SPM12 (Brett et al. 2002).

Psychophysiological Interactions. Lastly, we investigated the integration of social processing with vocal motor processing reflected in task-related connectivity changes during exerting social vocal control with the core vocal motor area in left IFG. To this end, we computed psychophysiological interactions (PPI) in SPM12. To isolate the individual core vocal motor area, we constructed VOIs (volumes of interest) based on the peak activation of the 2nd level group contrast of Go > No-Go trials in left IFG. A sphere of 10 mm radius was constructed around the peak of this activation map ([-52 12 24]) and used as a mask image to search for peak activation on a single-subject level with a voxel height threshold of $p <.001$. We then defined VOIs on the peak activation for each participant and built individual spheres with radius of 6 mm (i.e. 3 voxels) around this peak coordinate. The first eigenvariate of the functional MRI signal change was then extracted from the VOIs and the mean time course was multiplied by the task regressors. In separate PPI analyses, the task regressors were based on the following contrasts: 1) Go > No-Go for social modulation trials, and 2) affiliation (likeable and hostile) > competence (intelligence) trials. We added these interaction terms for each model as regressors to the 1st level models, along with the deconvolved source signal of the VOI and the task regressors. Contrasts between Go and No-Go trials for the relevant conditions (e.g. social vocal control) were chosen, because we were interested in exploring mechanisms specific to voice production, while subtracting out other general task-related effects that were common to both task types (e.g. imagining a social scene). Thus, the contrast of Go > No-Go allowed us to target the mechanisms specifically involved in achieving ongoing social voice production.

Statistical analysis of rating data

Vocal modulation performance. The ratings obtained from naïve listeners were analysed in R (http://www.R-project.org/). To assess the success of the voice manipulations to express social traits, we calculated the average change in naïve ratings for modulated voices relative to the normal voice samples, for the intended trait (e.g. comparing “intelligent” ratings for the normal and “intelligent” trials) and for the other traits (e.g. comparing “intelligent” ratings for the normal and the “likeable” trials) individually for each speaker and each trait. Thus, we obtained the change in mean ratings, henceforth Δ - ratings, from each speaker’s normal voice as a function of each expressed trait (e.g. intelligent ratings of “intelligent” modulation – intelligence ratings for the normal voice), and for each trait condition, allowing us to measure the effectiveness of volitional social trait expression, i.e. the sensitivity and specificity with which voice changes evoked the intended social trait percepts in listeners. The Δ - ratings were then analyzed in the framework of linear mixed effects models for each trait separately ('lme4' package; Bates, Mächler, Bolker, & Walker, 2015). Each model included the expressed trait as a fixed effect term and speaker as a random intercept to account for within subject variation. Likelihood ratio tests were performed to test the effect of trait expression on the Δ - ratings, by comparing the models with fixed effects to the null models with only the random intercepts. We implemented planned treatment contrasts with the congruent trait rating as a reference, to test directly whether sounds from the congruent trait condition received significantly higher trait ratings than all other voice modulation conditions. Statistical significance for all models was set at a Bonferroni corrected significance level of p =0.013 (for 4 comparisons within each model).

We also contrasted ratings on arousal and valence across the social trait modulations in the framework of linear mixed effects models, using the expressed trait as a fixed effect term and the speaker as a random intercept. Here, the mean ratings were entered into the model, because we aimed to statistically contrast differences in comparison to the normal voice. Therefore, we again implemented planned treatment contrasts in these models with the ratings obtained for the normal voice as reference. Again, statistical significance for all models was set at a Bonferroni corrected significance level of p =0.013 (for 4 comparisons within each model). Statistical assumptions for all implemented linear mixed effects models were tested and met.

Social trait space dimensions in voice modulation. To test the differentiation of the trait categories based on the multivariate naïve ratings (Δ - ratings, which measures the change in ratings relative to the normal voice on the different traits, as well as arousal and valence), we computed a linear discriminant analysis in R ('MASS' package; Venables & Ripley, 2002). The final model included the trait category (i.e. likeable, hostile, intelligent and large) as dependent variable and the mean – centered change in naïve ratings for the modulated voices from the normal voice (Δ - ratings) on all trait scales, as well as on arousal and valence, as the predictor variables. This allowed us to explore the contribution of individual trait and control (i.e. arousal and valence) ratings to discriminant functions and test whether these would be differentiable along the conceptual distribution of the affiliation and competence dimension of the social trait space (Fiske et al. 2007; McAleer et al. 2014; Belin et al. 2017). In other words, we tested whether changes in ratings evoked by modulating the voice represent exaggerations within the trait space.

Results

Behavioural results

Subjective ratings. Subjective performance (as assessed by the talkers themselves) differed among the social traits expressed in the social vocal control task, $\chi^2(4)=40.26$, $p<001$. Planned contrasts showed that likeable voice modulations (M=5.37, SD=.64) were perceived as equally successful as the normal (M=5.68, SD=0.72) voice expression, $t=-1.89$, $p=.06$. That is, normal voice trials sounded as normal as likeable modulations sounded likeable. All other trait modulations were perceived as less intensely expressed in the voice, yet above the midpoint of the 7-point Likert scale (hostile: M=5.20, SD=0.61, intelligent: M=4.59, SD=0.60, large: M=4.97, SD=0.76, all *t*s < -2.913, all *p*s<.01). This indicates that participants felt they were able to do the task, in spite of differences in performance between the traits. The subjective speaker rating for each recording was included as a parametric modulator in the fMRI analysis for the respective condition regressor to account for differences in functional activation related to task difficulty.

Vocal modulation performance. For analysis of the naïve listener ratings, all voices (normal and modulated) were rated on all trait scales (likeable, hostile, intelligent, large). There was a significant relative change in the naïve listener ratings evoked from social vocal modulation in all trait categories relative to normal voice recordings (Δ - ratings), indicating that social vocal modulation was perceivable by the listeners (all χ^2s(3)>29.40, *p*s<001, see Figure 2).

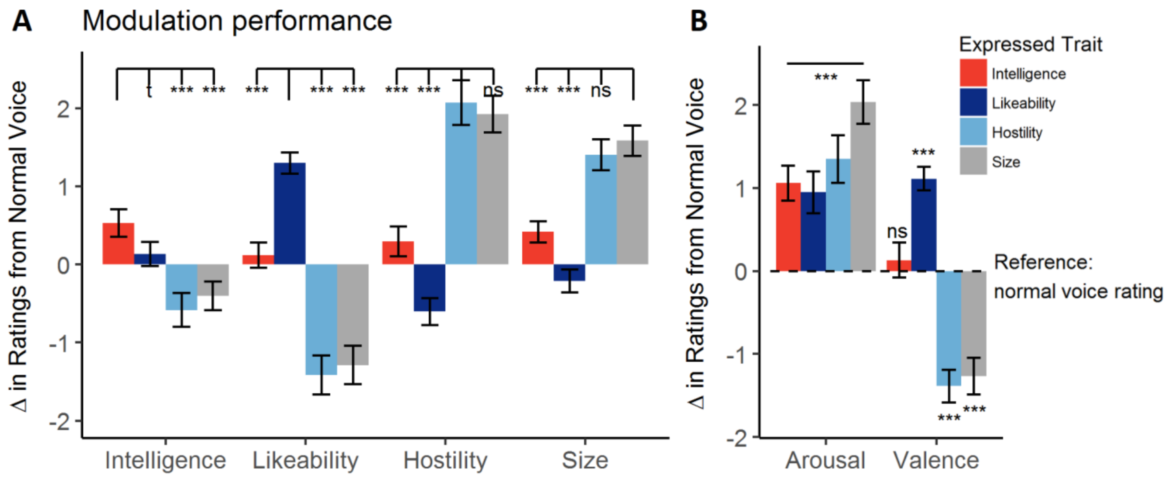


Figure 3. Results from univariate repeated-measures ANOVAs for each rated trait category. A. Changes in trait ratings relative to ratings of each speaker's normal voice (Δ-ratings) as a factor of social trait modulation. The models represent the comparison of congruent trait ratings to all incongruent trait ratings. **B**. Changes in intensity ratings on arousal and valence from normal voices as a factor of social trait modulation. ****p*<.001, ns = not significant. Contrasts were corrected for multiple comparisons using Bonferroni correction. Error bars = standard errors.

Planned contrasts showed that modulated voices expressing likeability were rated as sounding significantly more likeable than voices expressing other traits (all bs<-1.18, all ts(72)<-4.87, all ps<.001). Likeable voices were also perceived as more positive (b=-1.11, $t(96)$=5.29, p<.001) and higher in arousal than normal voices (b=0.90, $t(96)$=3.53, p<.001).

Hostile voice modulation lead to significantly greater relative ratings of hostility, than did likeable modulations (b=-2.68, $t(72)$=-10.11, p<.001), or intelligent voices (b=-1.76, $t(72)$=-6.71, p<.001). Voices expressing larger body-sizes were perceived as similarly hostile as hostile voices (b=-0.15, $t(72)$=-0.55, p=.58). Hostile voices were also rated as more negative (b=-1.39, $t(96)$=-6.60, p<.001) and higher in arousal than normal voices (b=1.35, $t(96)$=5.01, p<.001).

Intelligent voices were perceived as sounding more intelligent than hostile voices (b=-1.11, $t(72)$=-5.39, p<.001) or large voices (b=-0.93, $t(72)$=-4.51, p<.001). Likeable

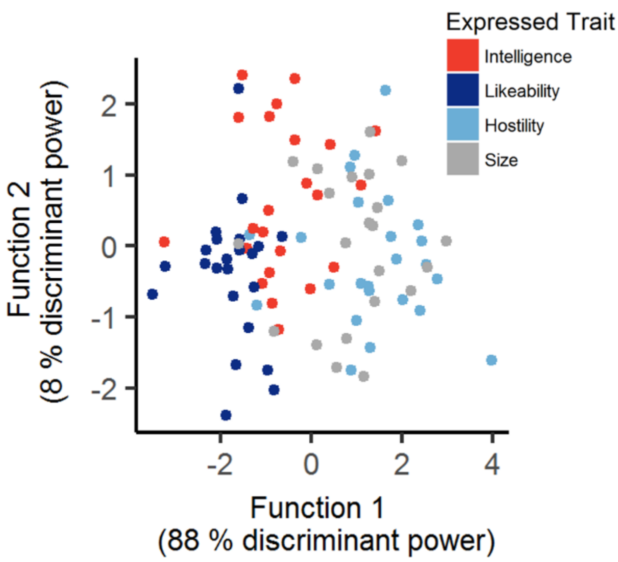


Figure 4. Distribution of linear discriminants of each recording resulting from the first two linear discriminant functions. Colours show the trait expressed by the speakers.

voices also gave relatively increased intelligence ratings, although marginally less so than intelligent voices (b=-0.40, t(72)=-1.92, p=.059). Intelligent voices were perceived as higher in arousal (b=1.06, t(96)=3.93, p<.001) but similarly neutral in valence as normal voices (all bs<0.22, all ts(96)<-1.13, all ps>.05).

Lastly, voices modulated to sound larger induced a positive change in size ratings that was significantly higher than for likeable voices (b=-1.80, t(72)=-8.60, p<.001) or intelligent voices (b=-1.17, t(72)=-5.59, p<.001). Large voices were rated similarly large as hostile voices (b=-0.18, t(72)=-0.88, p=.383), but were perceived as higher in arousal than hostile voices (b=2.04, t(96)=7.57, p<.001). Large voice modulations were also perceived as more negative than normal voices (b=-1.27, t(96)=-6.03, p<.001). Taken together, congruent trait ratings (i.e. when the expressed trait and the rated trait coincided) were generally rated significantly higher than incongruent trait ratings (i.e. when expressed traits did not coincide with trait ratings; see Figure 2).

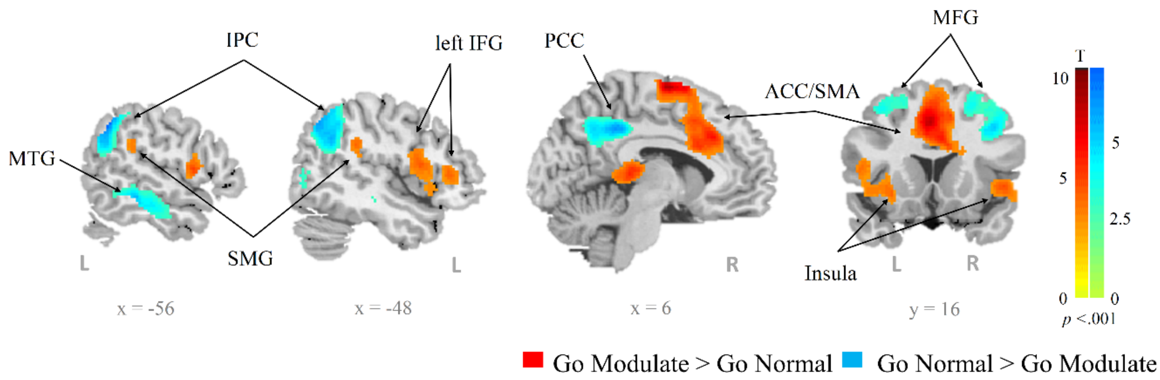


Figure 4. Vocal Modulation Network: Activation maps. The contrast Go Modulation > Go Normal voice (red) evoked changes in activation in ACC, IFG, Insula, SMA, SMG, STG. The contrast Go Normal voice > Go Modulation (blue) showed activation in IPC, MFG, MTG, Precuneus, PCC. ACC = anterior cingulate cortex, IFG = inferior frontal gyrus, IPC = inferior parietal cortex, MFG = middle frontal gyrus, MTG = middle temporal gyrus, SMA = supplementary motor area, SMG = supramarginal gyrus, STG = superior termporal gyrus. L= left, R = right.

Social trait space dimensions in voice modulation. Linear discriminant analysis further showed a successful differentiation between social voice modulations based on naïve ratings of traits and ratings of arousal and valence (Wilk's lambda=.355, $F(12,227.8)=7.06$, $p<.001$). Four recordings were removed as multivariate outliers from the analysis, leaving 92 modulation recordings in the model. The model showed an overall classification accuracy of 57% (95% CI = 0.46, 0.67), which was significantly above the No Information Rate (NIR = 26%, $p<.001$). Based on the ratings, likeable voices were best differentiated, with high sensitivity (87%) and specificity (83%), whereas large voices (from the body size condition) were least classifiable with a sensitivity of 32% but specificity of 87%. A combination of three linear discriminant functions allowed this classification, whereby the first two linear discriminants explained 96% of total variance. Based on the modest discriminant power of the third function, it was not analyzed further. The first function differentiated best between likeable and hostile voice modulation, accounting for 88% of explained between-group

Table 1. Functional activations for the contrasts go modulate > go normal voice and go normal > go modulate.

Contrast	*k*	Region	Hem.	Coordinate *x*	*y*	*z*	*T*	*Z*
Modulation > normal voice (go trials)	4817	Supplementary motor area	L	0	-4	68	10.55	6.31
		Pre-Supplementary motor area	L	-8	-2	72	9.28	5.93
		Anterior cingulate gyrus	L	-4	16	42	8.58	5.69
	1290	Inferior frontal gyrus, pars triangularis	L	-38	26	2	6.19	4.7
		Inferior frontal gyrus, pars opercularis	L	-56	8	10	5.62	4.41
		Insula	L	-32	18	-4	5.54	4.37
	374	Insula	R	48	12	-2	5.87	4.54
		Superior temporal gyrus	R	56	8	-4	4.82	3.97
			R	62	2	-2	4.7	3.9
	142	Supramarginal gyrus	L	-52	-36	28	4.9	4.01
	865	Corpus callosum, posterior part		0	-28	12	7.5	5.28
			R	12	-36	14	5.23	4.2
			R	4	-22	24	3.79	3.31
Normal voice > Modulation (go trials)	1215	Inferior parietal cortex	L	-54	-52	46	9.13	5.88
			L	-54	-56	32	7.46	5.27
	2530	Inferior parietal cortex	R	54	-54	34	8.92	5.81
			R	56	-48	46	8.06	5.50
			R	48	-60	36	7.99	5.47
	1306	Posterior cingulate cortex	R	6	-42	40	7.57	5.31
			R	2	-58	42	6.36	4.78
	2073	Superior frontal gyrus	R	26	30	46	7.54	5.30
		Middle frontal gyrus	R	40	18	40	6.58	4.89
			R	40	28	46	6.31	4.76
	816	Middle frontal gyrus	L	-34	24	48	6.59	4.89
			L	-38	30	44	6.55	4.87
		Superior frontal gyrus	L	-22	28	50	5.29	4.24
	554	Middle temporal gyrus	L	-56	-34	-10	6.79	4.98
			L	-62	-36	-4	6.4	4.8
			L	-58	-24	-16	5.84	4.53
	238	Middle temporal gyrus	R	62	-22	-14	5.73	4.47
			R	58	-6	-24	4.4	3.71
	108	Cerebellum	L	-38	-74	-34	5.13	4.14
	92	Orbital frontal gyrus	R	42	54	-6	5.10	4.13
		Middle frontal gyrus	R	46	54	4	4.0	3.45

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. $p<.001$ uncorrected, minimal cluster size: 61 voxels.

variance. Differentiation based on this discriminant function relied on hostile (*b*=.33), body-size (*b*=.50) and likeable (*b*=-.19) voice ratings, but not intelligence ratings (*b*=.02). Among the control ratings, valence ratings (*b*=-.45) and arousal (*b*=-.27) also contributed to the differentiation. The second discriminant differentiated between intelligent voices and voices expressing affiliation traits (likeable and hostile voices), accounting for 8% of the between-group variance (see Figure 3). Intelligence (b=1.14) ratings loaded on this discriminant in the opposite direction of hostile (b=-.32), likeable (b=-1.03), as well as body-size ratings (b=-.59). Arousal (b=.18) ratings also contributed to this function, whereas valence did not (b=.02). Thus, the changes in rating behaviour induced by the vocal modulation can be differentiated relative to the expressed trait and together, reflect the two dimensions of the social trait space affiliation and competence as reported previously (e.g. Fiske et al. 2007; McAleer et al. 2014; see Figure 3).

fMRI results

Group contrasts. Voice Modulation as opposed to normal voice (all modulation conditions > normal voice (Go trials)) induced changes in functional activation in 5 clusters, showing peak activations in the bilateral insulae, right superior temporal gyrus, left inferior frontal gyrus (IFG: triangular and opercular portions), supplementary motor area (SMA) and the anterior cingulate cortex (ACC), left supramarginal gyrus (SMG) and posterior parts of corpus callosum at a threshold of p<.001, minimal cluster size of k=61 voxels. In contrast, speaking in a normal voice elicited changes in activation in bilateral inferior parietal cortex (IPC), bilateral middle frontal gyurs (MFG), bilateral middle temporal gyrus (MTG), posterior cingulate cortex (PCC), and left cerebellum (see Figure 4 and Table 1).

Social voice modulations compared to nonsocial voice modulations (likeable ∩ hostile ∩ intelligent) > body-size, thresholded at uncorrected p<.001, minimum cluster size of k=61) induced changes in functional activation in 9 clusters with peak activation in left

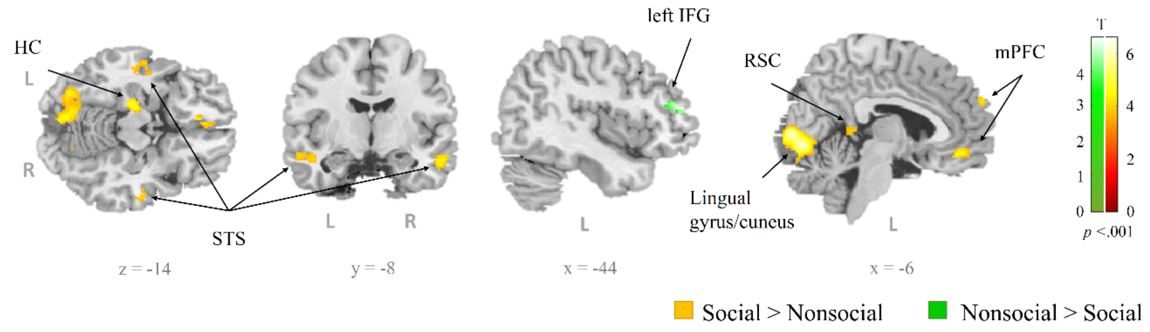


Figure 5. Social voice modulation: Activation maps. The contrast Social > Non-social voice modulation (yellow) activated areas including the bilateral STS, mPFC, left HC cortex, RSC, lingual gyrus, cuneus and precuneus (not depicted). Nonsocial > social voice change (green) lead to changes in activation in left triangular parts of the IFG. IFG = inferior frontal gyrus, mPFC = medial prefrontal cortex, HC = Hippocampus, RSC = retrospenial cortex, STS = superior temporal sulcus. L= left, R = right.

Table 2. Functional activations for the contrasts social > nonsocial modulation and nonsocial > social modulation.

Contrast	***k***	**Region**	**Hem.**	**Coordinate**			***T***	***Z***
				x	*y*	*z*		
Social > Non-social voice change (go trials)	1930	Lingual Gyrus	L	-8	-78	-6	6.61	4.90
				-10	-84	0	6.43	4.81
				0	-88	0	5.77	4.49
	86	Precuneus	L	-20	-48	36	5.02	4.08
	62	Cuneus	L	-20	-92	20	4.17	3.56
				-22	-86	32	3.83	3.33
	105	Hippocampus	L	-22	-20	-14	4.85	3.98
	111	Ventral medial prefrontal cortex	L	-6	44	-12	4.68	3.88
				-12	48	-6	3.86	3.36
	69	Dorsal medial prefrontal cortex	L	-4	62	28	4.45	3.45
	122	Superior temporal sulcus	L	-58	-8	-16	4.17	3.56
				-48	-6	-18	4.15	3.55
				-54	-18	-16	4.02	3.46
	144	Superior temporal sulcus	R	52	-8	-18	4.63	3.85
				58	-14	-14	4.18	3.57
	68	Retrosplenial cortex	L	-6	-40	6	3.9	3.38
			R	2	-46	12	3.81	3.32
Nonsocial > Social voice change (go trials)	65	Inferior frontal gyrus, pars triangularis	L	-44	46	16	4.96	4.05

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. *p*<.001 uncorrected, minimal cluster size: 61 voxels.

hippocampus (HC), dorsal and ventral portions of the medial prefrontal cortex (mPFC), left cuneus and precuneus, bilateral lingual gyri, bilateral superior temporal sulci (STS) and bilateral retrosplenial cortex. In contrast, nonsocial voice modulations (body-size > likeable ∩ hostile ∩ intelligent) compared to social voice changes engaged left-lateralized regions in the triangular portions of the IFG (see Figure 5 and Table 2).

During vocal modulation along the affiliation dimension of the social trait space (hostile ∩ likeable > intelligent, thresholded at $p<.001$, minimum cluster size of $k=58$), we found functional activation changes in the left amygdala, the right posterior STS/temporo-parietal junction (TPJ), right SMG, right precentral gyrus and an activation cluster spanning from posterior cingulate cortex to the precuneus. The competence dimension (intelligent > hostile ∩ likeable) induced activation changes in 4 clusters, including a cluster in the left IFG, spanning opercular and triangular portions and into medial frontal gyrus, inferior frontal gyrus pars orbitalis, bilateral superior frontal gyrus, including SMA, as well as a cluster spanning over lingual gyrus bilaterally and left cuneus/calcarine gyrus (see Figure 6 and Table 3).

PPI. The PPI analysis revealed changes in functional connectivity of the left IFG with regions in dorsal mPFC, right putamen, left posterior cingulate cortex (PCC) and precuneus, middle cingulate cortex, right posterior insula, right IFG (opercula and triangular portions) and cerebellum during social voice modulation trials (Go > No-Go social voice modulation). Inversely, during No-Go social modulation trials, functional connectivity increased between IFG and left anterior insula (see Figure 7 and Table 4). During trials requiring vocal modulation along the affiliation dimension, we observed changes in functional connectivity between the IFG and clusters in the dorsal striatum (spanning to anterior insula), middle and posterior portions of right STS, and right triangular and orbital portions of the IFG (see Figure 7 and Table 4). There was no meaningful change in functional connectivity between

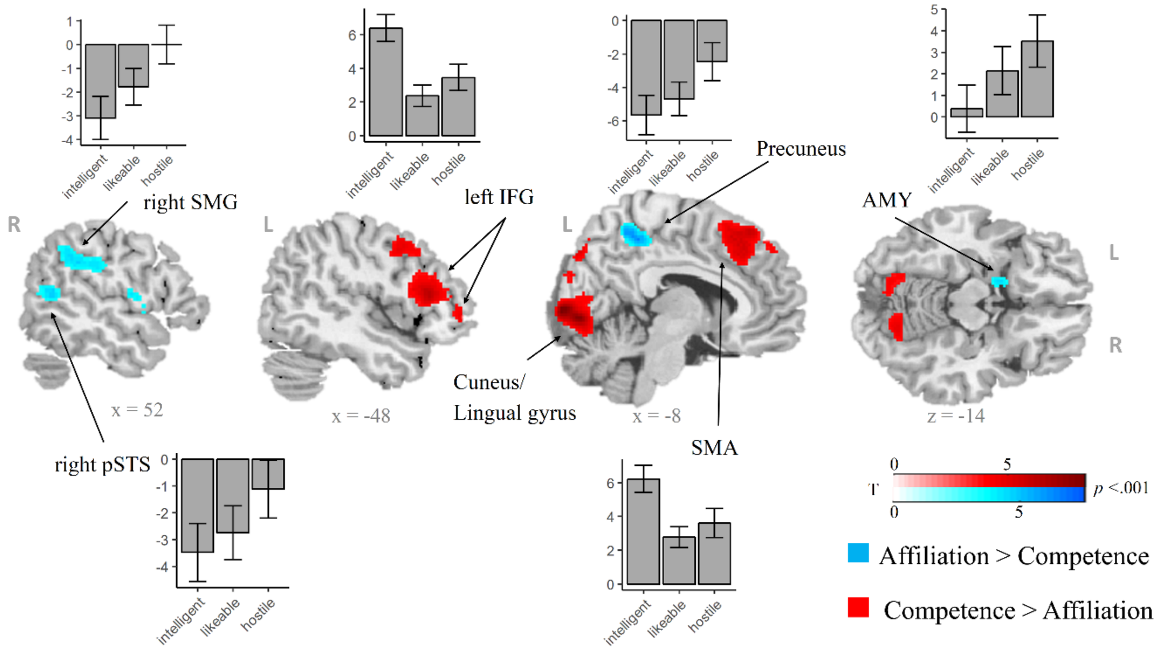


Figure 6. Social trait space modulation: Activation maps and parameter estimates. The contrast Affiliation > Competence (blue) induced changes in activation in clusters including right TPJ, right SMG, precuneus, bilateral AMY. The contrast Competence > Affiliation (red) evoked BOLD changes in left IFG pars opercularis and pars triangularis, ACC, SMA, cuneus and lingual gyrus. Parameter estimates illustrate evoked changes in response to each modulation condition. ACC = anterior cingulate cortex, AMY = amygdala, IFG = inferior frontal gyrus, mPFC = medial prefrontal cortex, HC = Hippocampus, PCC = posterior cingulate cortex, SMA = supplementary motor area, SMG = supramarginal gyrus. Bar plots illustrate parameter estimates (arbitrary units) in the significant cluster per condition compared to rest, error bars = standard errors. L= left, R = right.

the left IFG and other brain regions during competence voice modulation (see also supplementary materials S3). All contrasts were thresholded at uncorrected $p<.001$, minimal cluster size of 58 voxels.

Discussion

The voice is both a dynamic social behaviour and a rich source of information about a person. Successful modulation of the voice to express socially relevant information is an important contributor to achieving interactional and communicative goals. In this study, we found

Table 3. Functional activations for the contrasts Affiliation > Competence and Competence > Affiliation modulation.

Contrast	*k*	Region	Hem.	Coordinate *x*	*y*	*z*	*T*	*Z*
Affiliation > Competence	912	Precuneus	L	-12	-42	50	7.47	5.27
			R	2	-42	48	5.08	4.12
		Posterior cingulate gyurs	R	12	-34	44	4.77	3.94
	711	Supramarginal gyrus	R	60	-44	38	5.64	4.42
				58	-36	32	4.97	4.05
				48	-28	32	4.43	3.73
	219	Posterior superior temporal sulcus / Temporo-parietal junction	R	46	-54	14	5.29	4.24
				56	-58	10	5.02	4.08
	97	Amygdala	L	-18	4	-10	5.27	4.23
	74	Precentral gyrus	R	54	-2	10	4.24	3.60
Competence > Affiliation	2526	Cuneus	L	-10	-90	2	8.28	5.58
		Lingual gyrus	L	-8	-80	-2	7.79	5.40
		Calcarine gyrus		0	-84	-2	7.61	5.33
	2026	Superior frontal gyrus	L	-26	20	56	6.61	4.9
				-10	34	54	6.05	4.63
				0	26	46	5.33	4.26
	1218	Inferior frontal gyrus, pars triangularis	L	-48	26	12	6.0	4.61
		Middle frontal gyrus	L	-42	48	0	5.65	4.43
				-34	58	12	5.20	4.19
	67	Inferior frontal gyrus, pars orbitalis	L	-36	30	-6	4.81	3.96

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. $p<.001$ uncorrected, minimal cluster size: 58 voxels.

support for our hypotheses: First, we showed that social trait judgements are modulated along the social trait space dimensions as a function of social trait expression in voices. Second, neural activation data showed activation of processing networks related to social trait processing (STS, pSTS/TPJ, mPFC, Precuneus) and vocomotor control (left IFG, SMA, SMG, ACC) during the performance of voluntary, socially-relevant vocal modulation. Finally, functional connectivity analyses suggest an interaction between the left IFG and the social brain network during the performance of social vocal modulations.

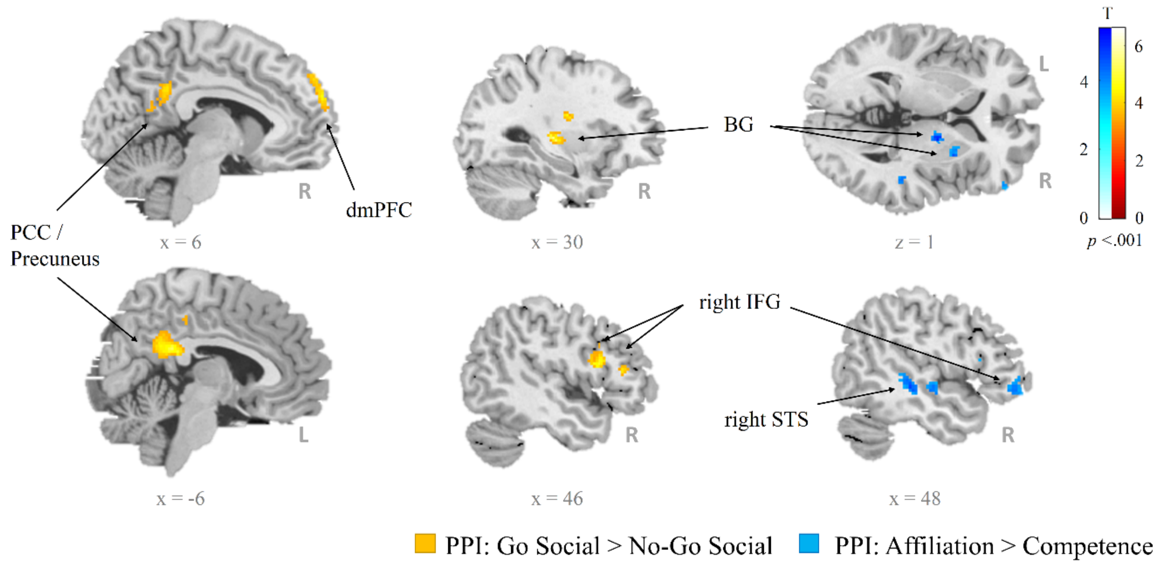


Figure 7. **PPI analysis connectivity maps, seed: left IFG**. Go Social > No-Go Social Voice Modulation (yellow). During social voice modulation left IFG was functionally connected to regions including the medial prefrontal regions, posterior cingulate cortex, BG and right IFG. Affiliation > Competence modulation (blue). Connectivity with the left IFG increased during social voice modulation along the affiliation continuum with right STS and IFG, as well as two clusters in the basal ganglia including (dorsal striatum). Seed: left IFG. BG = basal ganglia, dmPFC = dorso-medial prefrontal cortex, IFG = inferior frontal gyrus, PCC = posterior cingulate cortex. L= left, R = right.

Vocal control of social traits expression

Our behavioural data showed that volitional expression of social traits in the voice was not only recognizable as a voice change, but effective in its intention: it led to specific changes in perceptual ratings relevant to the targeted social trait. One previous study has investigated vocal modulation to express trait information (Hughes et al. 2014). In this study, the modulated speech exemplars were sequences of numbers, which were rated by naïve listeners and compared to the rating of the normal voice for the congruent social trait only. We extend these findings, showing evidence for specificity and effectiveness of vocal modulations. Further, we show that social voice modulations during speaking of pseudowords are differentiable for naïve listeners on a multivariate level, and that the change in rating

Table 4. Results of the PPI analysis.

Contrast Seed: L IFG	*k*	Region	Hem.	Coordinate *x*	*y*	*z*	*T*	*Z*
Go > No Go (social modulation)	718	Posterior cingulate cortex	L	-10	-42	22	6.81	4.99
				-4	-50	28	4.61	3.84
			R	4	-46	32	4.52	3.79
	60	Middle cingulate cortex	L	-2	-30	44	4.42	3.72
	103	Dorsal striatum	R	30	-22	4	5.67	4.44
	67	Posterior Insula	R	28	-6	22	5.02	4.09
				32	-14	20	4.49	3.77
	208	Dorsal medial prefrontal cortex	R	6	62	28	4.67	3.87
				4	58	38	4.14	3.54
				20	62	26	3.82	3.33
	283	Inferior frontal gyrus, pars triangularis	R	50	16	20	5.19	4.18
				48	30	14	4.37	3.69
		Inferior frontal gyrus, pars opercularis	R	50	8	22	4.87	4.0
	140	Cerebellum	L	-26	-76	-32	4.77	3.87
				-32	-68	-34	4.2	3.39
				-40	-60	-36	3.82	3.33
No Go > Go (social modulation)	79	Anterior Insula	L	-26	26	6	5.18	4.18
Affiliation > Competence	165	Dorsal Striatum	R	16	-2	4	5.49	4.35
			R	28	6	2	4.47	3.75
		Anterior Insula	R	30	24	14	4.33	3.74
	160	Superior temporal sulcus	R	52	-14	4	4.77	3.93
				50	-28	-4	4.68	3.88
				46	-34	0	4.44	3.74
	65	Inferior frontal gyrus, pars triangularis	R	44	22	10	4.69	3.89
				38	28	10	3.70	3.25
	102	Middle frontal gyrus	R	36	12	36	4.66	3.87
				28	14	46	4.24	3.61
	88	Inferior frontal gyrus, pars orbitalis	R	44	44	-10	4.26	3.62
				50	44	2	3.82	3.33
				30	44	-8	3.75	3.28

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. p<.001 uncorrected, minimal cluster size: 58 voxels.

behaviour relative to hearing normal voices was best discriminated on two discriminants relating to the affiliation and competence dimensions of the social trait space. Our data therefore suggest that voluntary social vocal modulations evoke a change in perception of the speaker's voice, which can amplify a social trait rating relative to their normal voice.

We observed that vocal modulations to increase perceived body size and hostility were most often perceptually confused compared to the other modulation conditions. This might be driven by the fact that acoustic information related to body size, the body size projection, is an important contributor to perceptions of emotions in voices (Chuenwattanapranithi et al. 2009). Anger in particular, is related to an increased body size projection through vocal tract elongation. Although in this work, we did not ask speakers to express any emotion, we acknowledge that expressing hostility and anger might have similar social implication, i.e. to keep a safe distance from a speaker. Nevertheless speakers reported very differentiable scenarios during the hostile compared to the large voice modulations (e.g. “speaking louder” when trying to increase body-size versus “speaking to somebody I don’t like” during expression of vocal hostility). This differentiation was observable in speakers on the neural level, where only the supplementary motor area (SMA) showed overlapping in functional engagement during the two conditions (see supplement S4). This implies differentiable underlying social cognitive processes when modulating the voice to express a large body size versus hostility, in the face of perceptual confusion of body-size and hostility on the receiver end.

Trait judgements in the social space have been replicated extensively (reviewed by Fiske et al., 2007) and might reflect domain-general processing of information about the intentions of others, including judgements of personality from voices (McAleer et al. 2014; Baus et al. 2019), faces (Todorov, Said, et al. 2008) or stereotyping of social groups (Fiske et al. 2002). Although the voice has predominately been studied as a vehicle for spoken language, theorists have described speech as actions primarily carrying intentional force (Austin 1975; Grice 1975). In fact, the voice is primarily a tool for interpersonal communication – vocal behaviour mainly occurs when there is an intention to communicate information to others. Thus, the voice is an important source of information about a speaker’s intentions. Our data underline the potential social benefit that could arise from successful vocal modulations in interactions with others, whether to express liking, maintain distance, or convey competence.

Proficiency in the volitional control of social expression in the voice might be an important contributor to successfully managing impressions in a variety of social situations, from job interviews (Schroeder and Epley 2015) to political campaigns (Pavela Banai et al. 2017). As such, it could be a strategic tool with which social opportunists might manage another's impressions to achieve beneficial outcomes. One previous study has explored the effect of affective subcomponents of the psychopathic personality on prosody perception networks (Aziz-Zadeh et al. 2010). This study suggested a positive association between affective empathy levels and functional activation in common motor and perception regions during listening to affective prosody. However, whilst showing decreased sensitivity to socio-affective cues (Blair et al. 2007), individuals scoring high on psychopathy are particularly effective in volitional affective facial expressions (Book et al. 2015) and unimpaired in social-cognitive tasks (Theory of Mind abilities; Winter, Spengler, Bermpohl, Singer, & Kanske, 2017). Whether this extends to the voice - that is, whether trait cognitive versus affective empathy also has an effect on social voice modulation efficiency - remains unclear. Thus, the potential dissociation between perceptual sensitivity to spontaneous social cues as compared to the production of volitional social cues might be an important avenue for future studies.

The VMN in voluntary voice modulation

We report a modulation network engaged in voluntary vocal control in left IFG, ACC, SMA, SMG, STG and insula. This network overlaps largely with the previously reported vocomotor network (VMN) involved in voluntary vocal control (left IFG, SMA, SMG, ACC, STC and insula; Barrett et al., 2004; Golestani & Pallier, 2007; McGettigan et al., 2013; Peschke, Ziegler, Eisenberger, & Baumgaertner, 2012; Pisanski, Cartei, et al., 2016; Reiterer, Hu, Sumathi, & Singh, 2013; Simmonds, Leech, Iverson, & Wise, 2014). Our experimental set-up allowed us to compare specifically which parts of the VMN are centrally engaged in non-social aspects of vocal control, by asking participants to modulate the expression of body size. Here, particularly left IFG (pars triangularis), showed significant changes in activation when speakers produced vocal modulation without social content. The left IFG is a central structure of the volitional vocal control network involved in all aspects of vocalization (laryngeal, orofacial and respiratory control; reviewed by Simonyan & Horwitz, 2011). It is anatomically connected to ACC and the laryngeal and facial motor areas in the primary motor and ventral premotor cortex. Representing sound maps for feed-forward processing (Tourville and Guenther 2011), as well as for motor coordination, l-IFG exerts top-down inhibitory control of the subcortical affective vocal network via the ACC (Hage and Nieder 2016). We corroborate the central role of left IFG in voluntary vocal control. Moreover, vocal body-size modulations are typically associated with changes of both fundamental frequency and formant frequencies (Pisanski, Mora, et al. 2016). The triangular part of the l-IFG shows enhanced activation during modulation of prosodic information of the voice (Agnew et al. 2017) and is involved in semantic processing of vocal information during speech perception (Gough et al. 2005). Keeping with these findings, our findings support the notion that l-IFG (in particular pars triangularis) provides input into the formation of sound maps for speech that encompass supra-segmental and prosodic vocal meaning.

Neural mechanisms underlying social voice modulation

Our social vocal control task required participants to modulate their voice to express social traits. We hypothesised that this task engages the SBN, a set of regions related to social cognitive functions, including social trait and self-referential processing. In line with our hypothesis, the social vocal modulation conditions (hostile, likeable, intelligent) engaged the dorsal and ventral portions of the medial prefrontal cortex (mPFC), the bilateral superior temporal sulci (STS), left hippocampal formation and precuneus more strongly than the non-social vocal modulation condition (body size). These areas comprise the SBN (Van Overwalle 2009; Schurz et al. 2014), and have been partly implicated in previous studies requiring socially meaningful voice production, during impersonations (McGettigan et al. 2013; Brown et al. 2019) or while volitionally modulating the voice within a social contexts (Klasen et al. 2018). In the current study, we show the first evidence for engagement of social processing areas during voluntary voice change to express beneficial social traits.

MPFC is a core SBN area involved in domain-general social cognitive processing (Bzdok et al., 2012; Molenberghs, Johnson, Henry, & Mattingley, 2016; Schurz et al., 2014; for review see Van Overwalle, 2009) and mental state inference from vocal cues (Dricu and Frühholz 2016). Importantly, mPFC is suggested to represent social knowledge (Krueger et al. 2009), related to the social trait space (Van Overwalle 2009; Ma et al. 2011, 2012, 2014, 2016; Van Overwalle et al. 2016) in reference to others *and* the self (Nicolle et al. 2012; see also meta-analytic evidence of mPFC representation of the psychological self in Hu et al. 2016). Our findings corroborate the involvement of mPFC in processing related to social trait knowledge and importantly, extend previous work in showing its involvement in social evaluation, related to both others' and *own actions towards others*. In the present study, we manipulated the modulation of the voice to express social traits, which can be understood as a goal-directed social behaviour (Wolpert et al. 2003). Such voluntary social behaviour has been suggested to rely on mPFC engagement (Thornton et al. 2019), with specific roles for ventral and dorsal regions (Krueger et al. 2009). While dorsal mPFC is thought to represent social goal orientation and execution, ventral mPFC activation reflects self-relevant outcome expectations by modelling social behaviour (Nicolle et al. 2012) based on social trait knowledge (Ma et al. 2014). In support, we found both ventral and dorsal regions in mPFC are engaged during social voice change. Moreover, specifically dorsal mPFC was functionally connected to l-IFG during socially modulated speech, suggesting it may serve as an input region to l-IFG. We therefore suggest that engagement of mPFC in this study might reflect the speaker accessing social trait knowledge to inform the formation of specific trait related vocomotor maps to achieve goal-directed vocal adjustments.

Together with mPFC, middle parts of the STS showed enhanced activation during social vocal control conditions. The STC contains both emotion-sensitive (Kreifelts et al. 2009) and voice-sensitive (Belin et al. 2000) areas, and similarly to the fusiform gyrus for face processing, has been reported to be involved in voice identity processing (Belin and Zatorre 2003; Schall et al. 2014) and expression (McGettigan et al. 2013). While STG is involved in auditory prediction modelling during vocal control (e.g. Frühholz et al., 2015; conceptualized in the DIVA-Model of speech production by Tourville & Guenther, 2011), STS is reliably involved in inferring communicative intent from observed actions (e.g. Redcay et al., 2016; Schurz et al., 2014). Our current findings suggest that activity in the STS might not only be related to processing intention in others, but communicating one’s own social intentions to others, too.

Lastly, precuneus involvement is seen in tasks requiring social trait inference (Tavares et al. 2015; Ma et al. 2016; Van Overwalle et al. 2016) and self-referential processing (Hu et al. 2016). Interestingly, precuneus was the only classic social brain area to be more strongly engaged during normal compared to modulated voice production, supporting previous work suggesting the role of precuneus in self-referential processing (Cabanis et al. 2013). Nevertheless, more work is needed to understand the specific neural and cognitive mechanisms involved in active social behaviour, particularly in interactive settings that allow us to capture the entirety of social interactions, i.e. the action-feedback loop between two interlocutors.

Differentiation of social trait space dimensions in the voice. Although some evidence points to an integrated, common trait code processing in mPFC and precuneus (Tavares et al. 2015; Ma et al. 2016; Van Overwalle et al. 2016), these studies either did not directly compare the two trait dimensions (Ma et al. 2016) or showed differences in the valence of the dimensions (Van Overwalle et al. 2016). Our data suggest that vocal control to express these two components is associated with separable sites of activation and different functional connectivity patterns with vocomotor control networks. Whereas expression of the competence trait led to a stronger engagement of vocomotor control areas in IFG and SMA, affiliative modulations elicited greater activity in a right-dominant network of pSTS/TPJ, SMG, PCC and the left amygdala.

The affiliation dimension connotes positive or negative intentions towards others, whereas the competence dimension suggests potency or power to act on such intentions (Fiske et al. 2007). We found that posterior portions of the STS/TPJ were engaged during expression of affiliative information in voices. Processing in the pSTS/TPJ region is particularly dedicated to evaluating others' immediate intentions (Saxe & Powell, 2006; for meta-analytic results see: Molenberghs et al., 2016; Schurz et al., 2014; Van Overwalle, 2009) when they are socially significant (Redcay et al. 2010). Given that the affiliation dimension is of high social significance, functional processing in pSTS/TPJ might specifically subserve basic intention encoding of a speaker's own intentions to achieve volitional vocal expression. Keeping with this interpretation, another structure specifically engaged during affiliative vocal modulation was the amygdala, which has frequently been associated with processing behaviourally relevant and salient stimuli (Ewbank et al. 2009). Previous voice research has shown that the amygdala is reactive to affective content in voices (Frühholz and Grandjean 2013; Dricu and Frühholz 2016; Pannese et al. 2016), and is involved in regulating emotional vocal output behaviour (Pichon and Kell 2013; Frühholz et al. 2015). We found that amygdala activation was most pronounced during hostile vocal modulation along the affiliation dimension - this is in line with an affective processing account (LeDoux 2012), but also with previous work showing the amygdala's involvement in encoding stereotypical information along the affiliation dimension in faces (Engell et al. 2007; Todorov, Baron, et al. 2008; Todorov, Said, et al. 2008).

We observed that intelligent voice modulations were less distinctly perceived than the likeability or hostility expressions. In a perceptual study, a similar network of pSTS/TPJ, SMG, medial portions of STS, PCC, ventral mPFC and amygdala emerged when subjects heard clearly expressed vocal intentions as opposed to ambiguous recordings (Hellbernd and Sammler 2018). Ambiguous expressions engaged SMA, IFG and insula, regions that overlap partly with the competence-related activations we report. The increased engagement of motor-related regions in our study therefore might be due to effects of difficulty in formulating the vocomotor plan to sound "intelligent", leading to perceptually ambiguous expressions. In fact, the fMRI participants judged their own intelligent vocal modulations as less successful than other trait modulations (see supplement S2). Although we accounted for such differences in task difficulty by introducing parametric modulators based on the self-ratings in the statistical models, further studies manipulating social vocal control on both dimensions and reflecting both poles (i.e. decreased as well as increased intelligence) will be needed to differentiate whether the topographical activation differences observed in the current study reflect social or vocomotor processing differences.

Linking social information with motor planning

Our results investigated the interplay of the VMN and SBN in the support of social voice modulation, by examining task-related changes in the functional connectivity of the left IFG. The l-IFG is the central executor providing speech sound maps for voluntary vocalizations (Tourville and Guenther 2011). As a primary input region to the vocomotor network, IFG might serve as a hub for integrating social and vocal information, which is then used to create sound templates for vocomotor translation. We found that SBN regions emerged as being functionally coupled with l-IFG during socially motivated voice production - namely dorsal mPFC, precuneus, right IFG – as well as regions in the basal ganglia. Expressing affiliative information in the voice (speaking in a likeable or hostile voice) revealed additional functional connectivity with right STS. This suggests that during ongoing voice production, social cognitive computations work together with l-IFG to inform motor coordination of the vocal tract. Interestingly, only dorsal mPFC showed significant functional connectivity with l-IFG during social vocal control, but not ventral mPFC. Dorsal mPFC is structurally connected with premotor and somatosensory areas (Öngür et al. 2003), and has been associated with own choice execution (Nicolle et al. 2012) and representing goal oriented social schemata (Krueger et al. 2009). The basal ganglia are part of a subcortical network involved in emotional prosody production (Aziz-Zadeh et al. 2010; Laukka et al. 2011; Pichon and Kell 2013; Frühholz et al. 2015; Mitchell et al. 2016) and are thought to have regulatory functional connectivity with the amygdala, motor and auditory cortices during affective vocal control (Pichon and Kell 2013; Frühholz et al. 2015; Klaas et al. 2015). Specifically, ventral and dorsal striatum show distinct roles during emotional versus neutral prosody in motor control planning and executing motor plans, respectively (Pichon and Kell 2013). In this previous study, the dorsal striatum was functionally connected to hippocampus, amygdala and motor cortices during angry prosody production. We corroborate these findings showing increased functional connectivity between dorsal striatal regions and l-IFG specifically for ongoing affiliative vocal control and engagement during voice production trials specifically (Go vs. No-Go trials; see supplementary data S3 Table 1). This supports the involvement of the dorsal striatum in volitional vocal control (Laukka et al. 2011), but suggests that this involvement is not exclusively due to affect, but might be a more generalized function for socially relevant volitional vocal modulations.

Limitations. We acknowledge important differences between the experimental balancing of the two conditions: where both poles of the affiliation dimension were manipulated in the design, only one pole of the competence dimension was implemented. This is because we set out to test specifically the expression of *beneficial* social traits, in the sense that sounding likeable, hostile, and intelligent are helpful in achieving desired interactional outcomes. Whereas trying to come across as intelligent or creating distance or proximity to others may be beneficial, sounding deliberately unintelligent might not naturally lead to valuable outcome. Nevertheless, this imbalance might have impacted comparison between the two dimensions. In addition, although vocal modulations were effective in our study, we cannot account for individual differences in the strategies used to carry out the task. Speakers mainly reported imagining speaking to a known person towards whom they would have liked to express the traits presented, or to whom they had done so in the past. However, this might not be an analogue to how such processes would unfold in novel real-life interactions, and more studies are aiming to provide direct social interaction in fMRI settings to raise ecological validity (Schilbach 2016). Moreover, volitional voice modulation might present a very specific social situation which is rarely practised with such purity in everyday life. We suggest, however, that while voice manipulations may often be the result of spontaneous reactions, the extent of humans' flexible control over vocal expression allows for modulations to be strategically employed in interactions with others.

Lastly, we acknowledge that the majority of speakers were female (n=21). Few studies have shown sex-related differences in vocal modulation strategies during courtship, (e.g. Fraccaro et al. 2013; Pisanski et al. 2018). In this study, we were interested in the perceptual effects of vocally expressing socially desirable traits. In fact, relating to the social voice space, McAleer and colleagues (2014) report a consistent pattern in the perception of traits in listeners' ratings of male and female neutral voices (with the exception of perceived attractiveness). This supports the notion that despite the different acoustic modulation strategies that could be present in male and female speaking styles, they nonetheless lead to similar ends in terms of the impressions made on listeners. In fact, to our knowledge only one study has directly investigated volitional vocal control in males and females, suggesting similar efficacy in vocal modulation for most social traits, apart from expressions of confidence (Hughes et al. 2014). Future studies should nevertheless aim to obtain more balanced samples of gender identities to explore potential differences in social voice change and its neural mechanisms.

Conclusions and Outlook. A number of questions arise from our findings. Our results suggest that social brain areas work together with vocomotor control areas to achieve social vocal control. Although providing some first conclusive results, the underlying neural mechanisms remain unclear, such as which specific social processing functions underlie activation in social brain areas during this type of voice modulation. Moreover, future studies could investigate how individual differences in the efficacy of social voice modulation relate to different levels of social reactivity and mentalizing. For example, are more empathic individuals also better at encoding social information in their voice? Additionally, it remains to be determined how vocal modulation skills arise, as they comprise both social knowledge and fine-tuned motor control. Do we instinctively learn to express social information in the voice in the same way that we learn to speak, via our innate capacity as vocal learners? Lastly, we have introduced an intuitive vocal modulation task that requires targeted social evaluation and forecasting: this could be a candidate for theory of mind tasks of social expression that could be implemented in isolated and dialogic scenarios.

This study advances our understanding of the neural mechanisms involved in intentional vocal modulation during encoding of social trait information. We suggest that social vocal control can be exerted to reinforce percepts of traits across the social voice space and is therefore effective in conveying self-referential social intent. Our findings suggest that vocomotor control areas work together with social brain networks to achieve social vocal modulations, thereby extending previous work focussing on affective voice modulation. We suggest that precuneus and mPFC might be engaged in goal- and outcome-oriented self-referential trait processing, while STS activity might relate to intention encoding to achieve volitional social voice change. In sum, this study underlines the importance of the voice as a social behaviour and suggests that vocomotor networks interact with social processing streams to achieve dynamic vocal behaviours, with goal-directed social effects.

Funding. This work was supported by a Research Leadership Award from The Leverhulme Trust (RL-2016-013) awarded to CM.

Notes. We thank Elise Kanber for her assistance with data collection and participant training, as well as Ari Lingeswaran for technical help with MRI data collection. Conflict of Interest. None declared.

References

Ackermann H, Hage SR, Ziegler W. 2014. Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. Behav Brain Sci. 72:1–84.

Agnew ZK, Ward L, McGettigan C, Josephs O, Scott SK. 2017. Distinct neural systems for the production of communicative vocal behaviors. bioRxiv. 107441.

Austin JL. 1975. How to do things with words. 2nd ed. Cambridge, MA: Harvard University Press.

Aziz-Zadeh L, Sheng T, Gheytanchi A. 2010. Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. PLoS One. 5:e8759.

Banse R, Scherer KR. 1996. Acoustic profiles in vocal emotion expression. J Pers Soc Psychol. 70:614–636.

Barrett J, Pike GB, Paus T. 2004. The role of the anterior cingulate cortex in pitch variation during sad affect. Eur J Neurosci. 19:458–464.

Bates D, Mächler M, Bolker BM, Walker SC. 2015. Fitting linear mixed-effects models using lme4. J Stat Softw. 67:1–48.

Baus C, McAleer P, Marcoux K, Belin P, Costa A. 2019. Forming social impressions from voices in native and foreign languages. Sci Rep. 9:414.

Belin P, Boehme B, McAleer P. 2017. The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. PLoS One. 12:e0185651.

Belin P, Zatorre RJ. 2003. Adaptation to speaker's voice in right anterior temporal lobe. Neuroreport. 14:2105–2109.

Belin P, Zatorre RJ, Lafaille P, Ahad P, Pike B. 2000. Voice-selective areas in human auditory cortex. Nature. 403:309–312.

Belyk M, Brown S. 2016. Pitch underlies activation of the vocal system during affective vocalization. Soc Cogn Affect Neurosci. 11:1078–1088.

Blair KS, Smith BW, Mitchell DG V, Morton J, Vythilingam M, Pessoa L, Fridberg D, Zametkin A, Sturman D, Nelson EE, Drevets WC, Pine DS, Martin A, Blair RJR. 2007. Modulation of emotion by cognition and cognition by emotion. Neuroimage. 35:430–440.

Book A, Methot T, Gauthier N, Hosker-Field A, Forth A, Quinsey V, Molnar D. 2015. The mask of sanity revisited: Psychopathic traits and affective mimicry. Evol Psychol Sci. 1:91–102.

Brainard DH. 1997. The Psychophysics Toolbox. Spat Vis. 10:433–436.

Brett M, Anton J-LL, Valabregue R, Poline J-B. 2002. Region of interest analysis using an SPM toolbox. Abstract Presented at the 8th International Conference on Functional Mapping of the Human Brain, June 2-6, 2002, Sendai, Japan. Neuroimage. 16:2.

Brown S, Cockett P, Yuan Y. 2019. The neuroscience of Romeo and Juliet: An fMRI study of acting. R Soc Open Sci. 6:181908.

Bzdok D, Schilbach L, Vogeley K, Schneider K, Laird AR, Langner R, Eickhoff SB. 2012. Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. Brain Struct Funct. 217:783–796.

Cabanis M, Pyka M, Mehl S, Müller BW, Loos-Jankowiak S, Winterer G, Wölwer W, Musso F, Klingberg S, Rapp AM, Langohr K, Wiedemann G, Herrlich J, Walter H, Wagner M, Schnell K, Vogeley K, Kockler H, Shah NJ, Stöcker T, Thienel R, Pauly K, Krug A, Kircher T. 2013. The precuneus and the insula in self-attributional processes. Cogn Affect Behav Neurosci. 13:330–345.

Cartei V, Cowles W, Banerjee R, Reby D. 2014. Control of voice gender in pre-pubertal children. Br J Dev Psychol. 32:100–106.

Dricu M, Frühholz S. 2016. Perceiving emotional expressions in others: Activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. Neurosci Biobehav Rev.

Engell AD, Haxby J V., Todorov A. 2007. Implicit trustworthiness decisions: Automatic coding of face properties in the human amygdala. J Cogn Neurosci. 19:1508–1519.

Ewbank MP, Barnard PJ, Croucher CJ, Ramponi C, Calder AJ. 2009. The amygdala response to images with impact. Soc Cogn Affect Neurosci. 4:127–133.

Fiske ST, Cuddy AJC, Glick P. 2007. Universal dimensions of social cognition: warmth and competence. Trends Cogn Sci.

Fiske ST, Cuddy AJC, Glick P, Xu J. 2002. A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. J Pers Soc Psychol. 82:878–902.

Fraccaro PJ, O'Connor JJM, Re DE, Jones BC, DeBruine LM, Feinberg DR. 2013. Faking it: Deliberately altered voice pitch and vocal attractiveness. Anim Behav. 85:127–136.

Frühholz S, Grandjean D. 2013. Amygdala subregions differentially respond and rapidly adapt to threatening voices. Cortex. 49:1394–1403.

Frühholz S, Klaas HS, Patel S, Grandjean D. 2015. Talking in fury: The cortico-subcortical network underlying angry vocalizations. Cereb Cortex. 25:2752–2762.

Golestani N, Pallier C. 2007. Anatomical correlates of foreign speech sound production. Cereb Cortex. 17:929–934.

Gough PM, Nobre AC, Devlin JT. 2005. Dissociating linguistic processes in the left inferior frontal cortex with transcranial magnetic stimulation. J Neurosci. 25:8010–8016.

Grice HP. 1975. Logic and conversation. In: Cole P,, Morgan J, editors. Syntax and semantics. New York: Academic Press. p. 41–58.

Hage SR, Nieder A. 2016. Dual Neural Network Model for the Evolution of Speech and Language. Trends Neurosci.

Harris LT, Fiske ST. 2007. Social groups that elicit disgust are differentially processed in mPFC. Soc Cogn Affect Neurosci. 2:45–51.

Harris LT, Todorov A, Fiske ST. 2005. Attributions on the brain: Neuro-imaging dispositional inferences, beyond theory of mind. In: NeuroImage. p. 763–769.

Hellbernd N, Sammler D. 2016. Prosody conveys speaker's intentions: Acoustic cues for speech act perception. J Mem Lang. 88:70–86.

Hellbernd N, Sammler D. 2018. Neural bases of social communicative intentions in speech. Soc Cogn Affect Neurosci. 13:604–615.

Hu C, Di X, Eickhoff SB, Zhang M, Peng K, Guo H, Sui J. 2016. Distinct and common aspects of physical and psychological self-representation in the brain: A meta-analysis of self-bias in facial and self-referential judgements. Neurosci Biobehav Rev.

Hughes SM, Mogilski JK, Harrison MA. 2014. The Perception and Parameters of Intentional Voice Manipulation. J Nonverbal Behav. 38:107–127.

Jiang X, Pell MD. 2017. The sound of confidence and doubt. Speech Commun. 88:106–126.

Klaas HS, Frühholz S, Grandjean D. 2015. Aggressive vocal expressions-an investigation of their underlying neural network. Front Behav Neurosci. 9:121.

Klasen M, von Marschall C, Isman G, Zvyagintsev M, Gur RC, Mathiak K. 2018. Prosody production networks are modulated by sensory cues and social context. Soc Cogn Affect Neurosci. 13:418–429.

Ko SJ, Sadler MS, Galinsky AD. 2014. The Sound of Power: Conveying and Detecting

Hierarchical Rank Through Voice. Psychol Sci. 26:3–14.

Krauss RM, Freyberg R, Morsella E. 2002. Inferring speakers' physical attributes from their voices. J Exp Soc Psychol. 38:618–625.

Kreifelts B, Ethofer T, Shiozawa T, Grodd W, Wildgruber D. 2009. Cerebral representation of non-verbal emotional perception: fMRI reveals audiovisual integration area between voice- and face-sensitive regions in the superior temporal sulcus. Neuropsychologia. 47:3059–3066.

Krueger F, Barbey AK, Grafman J. 2009. The medial prefrontal cortex mediates social event knowledge. Trends Cogn Sci. 13:103–109.

Laukka P, Åhs F, Furmark T, Fredrikson M. 2011. Neurofunctional correlates of expressed vocal affect in social phobia. Cogn Affect Behav Neurosci. 11:413–425.

LeDoux J. 2012. Rethinking the Emotional Brain. Neuron.

Ma N, Baetens K, Vandekerckhove M, Kestemont J, Fias W, Van Overwalle F. 2014. Traits are represented in the medial prefrontal cortex: An fMRI adaptation study. Soc Cogn Affect Neurosci. 9:1185–1192.

Ma N, Baetens K, Vandekerckhove M, Van der Cruyssen L, Van Overwalle F. 2013. Dissociation of a trait and a valence representation in the mPFC. Soc Cogn Affect Neurosci. 9:1506–1514.

Ma N, Vandekerckhove M, Van Hoeck N, Van Overwalle F. 2012. Distinct recruitment of temporo-parietal junction and medial prefrontal cortex in behavior understanding and trait identification. Soc Neurosci. 7:591–605.

Ma N, Vandekerckhove M, van Overwalle F, Seurinck R, Fias W. 2011. Spontaneous and intentional trait inferences recruit a common mentalizing network to a different degree: Spontaneous inferences activate only its core areas. Soc Neurosci. 6:123–138.

Ma N, Wang S, Yang Q, Feng T, Van Overwalle F. 2016. The neural representation of competence traits: An fMRI study. Sci Rep. 6:39609.

McAleer P, Todorov A, Belin P. 2014. How do you say "hello"? Personality impressions from brief novel voices. PLoS One. 9:e90779.

McGettigan C. 2015. The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication. Front Hum Neurosci. 9:129.

McGettigan C, Eisner F, Agnew ZK, Manly T, Wisbey D, Scott SK. 2013. T'ain't what you say, it's the way that you say it--left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. J Cogn Neurosci. 25:1875–1886.

Mitchell RLC, Jazdzyk A, Stets M, Kotz SA. 2016. Recruitment of Language-, Emotion- and Speech-Timing Associated Brain Regions for Expressing Emotional Prosody: Investigation of Functional Neuroanatomy with fMRI. Front Hum Neurosci. 10:518.

Molenberghs P, Johnson H, Henry JD, Mattingley JB. 2016. Understanding the minds of others: A neuroimaging meta-analysis. Neurosci Biobehav Rev.

Morningstar M, Dirks MA, Huang S. 2017. Vocal Cues Underlying Youth and Adult Portrayals of Socio-emotional Expressions. J Nonverbal Behav. 41:155–183.

Nicolle A, Klein-Flügge MC, Hunt LT, Vlaev I, Dolan RJ, Behrens TEJ. 2012. An Agent Independent Axis for Executed and Modeled Choice in Medial Prefrontal Cortex. Neuron. 75:1114–1121.

Oleszkiewicz A, Pisanski K, Lachowicz-Tabaczek K, Sorokowska A. 2017. Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. Psychon Bull Rev. 24:856–862.

Öngür D, Ferry AT, Price JL. 2003. Architectonic subdivision of the human orbital and medial prefrontal cortex. J Comp Neurol. 460:425–449.

Pannese A, Grandjean D, Frühholz S. 2016. Amygdala and auditory cortex exhibit distinct sensitivity to relevant acoustic features of auditory emotions. Cortex. 85:116–125.

Pavela Banai I, Banai B, Bovan K. 2017. Vocal characteristics of presidential candidates can predict the outcome of actual elections. Evol Hum Behav. 38:309–314.

Peschke C, Ziegler W, Eisenberger J, Baumgaertner A. 2012. Phonological manipulation between speech perception and production activates a parieto-frontal circuit. Neuroimage. 59:788–799.

Pichon S, Kell CA. 2013. Affective and sensorimotor components of emotional prosody generation. J Neurosci. 33:1640–1650.

Pisanski K, Cartei V, McGettigan C, Raine J, Reby D. 2016. Voice Modulation: A Window into the Origins of Human Vocal Control? Trends Cogn Sci.

Pisanski K, Fouquet M, Mathevon N, Reby D. 2016. A longitudinal study of within-individual variation in human voice pitch. J Acoust Soc Am. 140:3397–3397.

Pisanski K, Mora EC, Pisanski A, Reby D, Sorokowski P, Frackowiak T, Feinberg DR. 2016.

Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. Sci Rep. 6:34389.

Pisanski K, Oleszkiewicz A, Plachetka J, Gmiterek M, Reby D. 2018. Voice pitch modulation in human mate choice. Proc R Soc B Biol Sci. 285:20181634.

Ponsot E, Burred JJ, Belin P, Aucouturier JJ. 2018. Cracking the social code of speech prosody using reverse correlation. Proc Natl Acad Sci U S A. 115:3972–3977.

Redcay E. 2008. The superior temporal sulcus performs a common function for social and speech perception: Implications for the emergence of autism. Neurosci Biobehav Rev.

Redcay E, Dodell-Feder D, Pearrow MJ, Mavros PL, Kleiner M, Gabrieli JDE, Saxe R. 2010. Live face-to-face interaction during fMRI: A new tool for social cognitive neuroscience. Neuroimage. 50:1639–1647.

Redcay E, Velnoskey KR, Rowe ML. 2016. Perceived communicative intent in gesture and language modulates the superior temporal sulcus. Hum Brain Mapp. 37:3444–3461.

Reiterer SM, Hu X, Sumathi TA, Singh NC. 2013. Are you a good mimic? Neuro-acoustic signatures for speech imitation ability. Front Psychol. 4:782.

Sauter DA, Eisner F, Calder AJ, Scott SK. 2010. Perceptual cues in nonverbal vocal expressions of emotion. Q J Exp Psychol. 63:2251–2272.

Saxe R, Powell LJ. 2006. Specific Brain Regions for One Component of Theory of Mind. Psychol Sci. 17:692–699.

Schall S, Kiebel SJ, Maess B, Von Kriegstein K. 2014. Voice identity recognition: Functional division of the right STS and its behavioral relevance. J Cogn Neurosci. 27:280–291.

Schilbach L. 2016. Towards a second-person neuropsychiatry. Philos Trans R Soc B Biol Sci.

Schroeder J, Epley N. 2015. The Sound of Intellect: Speech Reveals a Thoughtful Mind, Increasing a Job Candidate's Appeal. Psychol Sci. 26:877–891.

Schurz M, Radua J, Aichhorn M, Richlan F, Perner J. 2014. Fractionating theory of mind: A meta-analysis of functional brain imaging studies. Neurosci Biobehav Rev.

Shultz S, Vouloumanos A, Pelphrey K. 2012. The Superior Temporal Sulcus Differentiates Communicative and Noncommunicative Auditory Signals. J Cogn Neurosci. 24:1224–1232.

Simmonds AJ, Leech R, Iverson P, Wise RJS. 2014. The response of the anterior striatum during adult human vocal learning. J Neurophysiol. 112:792–801.

Simonyan K, Horwitz B. 2011. Laryngeal motor cortex and control of speech in humans.

Neuroscientist. 17:197–208.

Skoog Waller S, Eriksson M. 2016. Vocal Age Disguise: The Role of Fundamental Frequency and Speech Rate and Its Perceived Effects. Front Psychol. 7:1814.

Slotnick SD, Moo LR, Segal JB, Hart J. 2003. Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. Cogn Brain Res. 17:75–82.

Tavares RM, Mendelsohn A, Grossman Y, Williams CH, Shapiro M, Trope Y, Schiller D. 2015. A Map for Social Navigation in the Human Brain. Neuron. 87:231–243.

Thornton XMA, Weaverdyck ME, Tamir XI. 2019. The social brain automatically predicts others' future mental states. J Neurosci. 39:140–148.

Todorov A, Baron SG, Oosterhof NN. 2008. Evaluating face trustworthiness: A model based approach. Soc Cogn Affect Neurosci. 3:119–127.

Todorov A, Mandisodza AN, Goren A, Hall CC. 2005. Psychology: Inferences of competence from faces predict election outcomes. Science (80-). 308:1623–1626.

Todorov A, Said CP, Engell AD, Oosterhof NN. 2008. Understanding evaluation of faces on social dimensions. Trends Cogn Sci. 12:455–460.

Tourville JA, Guenther FH. 2011. The DIVA model: A neural theory of speech acquisition and production. Lang Cogn Process. 26:952–981.

Van Duynslaeger M, Van Overwalle F, Verstraeten E. 2007. Electrophysiological time course and brain areas of spontaneous and intentional trait inferences. Soc Cogn Affect Neurosci. 2:174–188.

Van Overwalle F. 2009. Social cognition and the brain: A meta-analysis. Hum Brain Mapp.

Van Overwalle F, Ma N, Baetens K. 2016. Nice or nerdy? The neural representation of social and competence traits. Soc Neurosci. 11:567–578.

Venables WN, Ripley BD. 2002. Modern Applied Statistics with S. 4th ed. New York: Springer. http://www.stats.ox.ac.uk/pub/MASS4.

Winter K, Spengler S, Bermpohl F, Singer T, Kanske P. 2017. Social cognition in aggressive offenders: Impaired empathy, but intact theory of mind. Sci Rep. 7:670.

Wolpert DM, Doya K, Kawato M. 2003. A unifying computational framework for motor control and social interaction. Philos Trans R Soc B Biol Sci.

Supplementary Materials

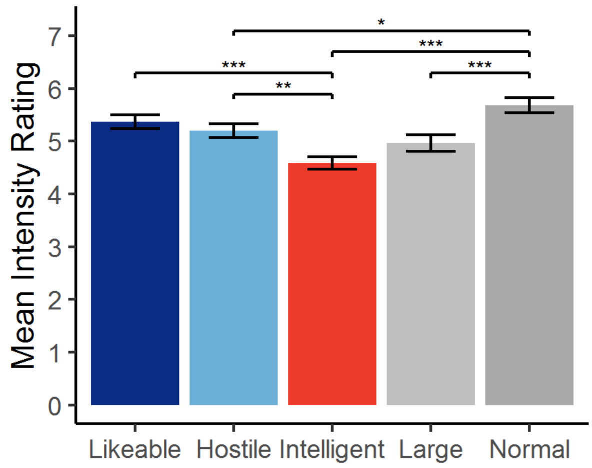
S1. Social trait scenarios

Participants received a short introduction to the social traits for the vocal modulation task prior to scanning. During this training session, all speakers were given the same instructions for each social trait, including ‘Normal: Speak in your normal voice and without trying to express anything in particular’, ‘Intelligent: Speak as if you were at a scholarly conference giving a presentation’, ‘Likeable: Speak as if you were trying to be liked by someone else, that you admire or try to make new friends`, ‘Hostile: Speak as if you want to distance yourself from someone, whom you strongly dislike or who has attacked you in the past’ and ‘Large: Speak as if you were trying to sound physically taller than you are’.

S2. Post-hoc contrasts: Subjective performance ratings

We obtained subjective trial-by-trial performance ratings (self-ratings) in a post-scanner session as reported in the main text. Here, speakers were presented with all in-scanner recordings and asked to rate how intensely they had managed to express the target trait on a 7-point Likert scale ranging from (1 not at all, and 7 very much). We conducted additional post-hoc Tukey contrasts to explore how the trait ratings differed from each other. These contrasts showed that speakers felt that they portrayed intelligence least intensely (M=4.59), showing a significant mean difference from hostility (M=-0.61, p<.01), likeability (M=-0.78, p<.001) and their normal voice (M=-1.1, p<.001). Hostility (M=-0.48, p<.05) and body size (M=-0.72, p<.001) were also expressed less intensely compared to normal voice. All post-hoc comparisons were Bonferroni corrected (see S2 Figure 1 below).

S2 Figure 1. Bar plot of speakers' subjective mean intensity ratings for each expressed social trait and results from post-hoc contrasts. **p*<0.05, ***p*<0.01, ****p*<0.001, error bars = standard errors.



S3. Psychophysiological Interaction Analysis

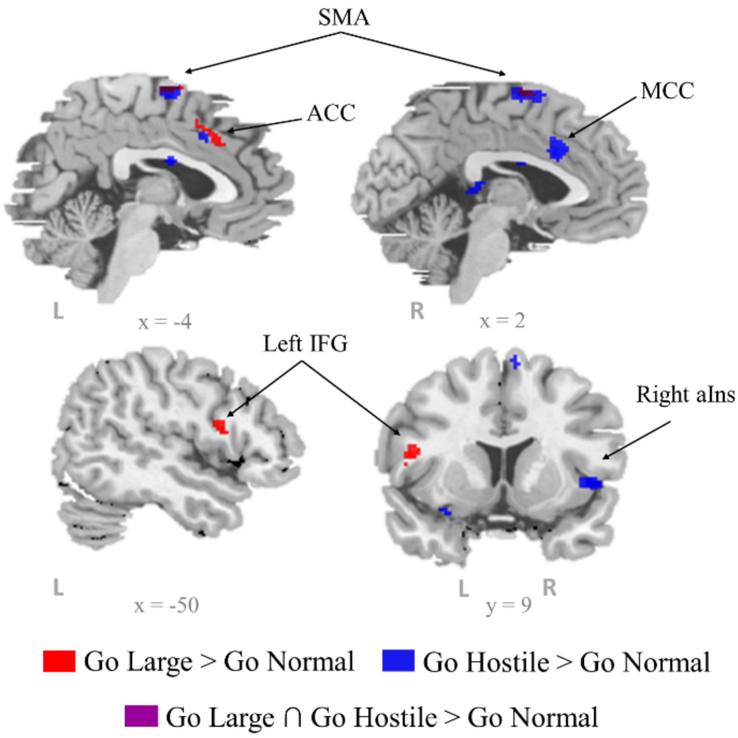
S3 Table 1. Functional connectivity analysis during Go trials of social trait space voice modulations. Results of the PPI analysis of contrasts Go Affiliation > No-Go Affiliation. No voxels survived the contrasts Go Competence > No-Go Competence or No-Go > Go Competence modulation.

Contrast Seed: L IFG	*k*	Region	Hem.	Coordinate *x*	*y*	*z*	*T*	*Z*
Go > No Go (Affiliation)	1812	Posterior cingulate cortex	L	-10	-44	20	6.34	4.78
				-4	-48	2	5.19	4.18
			R	4	-46	28	5.02	4.08
	832	Medial prefrontal cortex	R	4	58	34	5.37	4.28
			L	-4	50	44	4.98	4.06
		Supplementary motor area	L	-10	20	66	4.96	4.05
	224	Inferior frontal gyrus pars opercularis	R	48	16	22	5.07	4.11
				44	14	30	4.41	3.71
		Inferior frontal gyrus pars triangularis	R	56	28	16	4.27	3.63
	58	Dorsal striatum	R	26	-20	4	4.47	3.75
				20	-16	8	3.64	3.20
No Go > Go (Affiliation)	77	Anterior Insula	R	30	30	10	4.39	3.70
				28	22	8	4.04	3.47

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. *p*<.001 uncorrected, minimal cluster size: 58 voxels.

S4. Differential functional neural involvement during body-size and hostile voice modulations.

S4 Figure 1. Activation maps for the contrast Go Large > Go Normal (red), Go Hostile > Go Normal (blue) and the overlap between Go Large ∩ Go Hostile > Go Normal (purple). Overlapping activation is located in SMA. ACC = anterior cingulate cortex, aIns = anterior Insula, IFG = inferior frontal gyrus, MCC = middle cingulate cortex, SMA = supplementary motor cortex. L= left, R = right.



S4 Table 1. Functional activations for the contrasts go large > go normal voice and go hostile > go normal.

Contrast	***k***	**Region**	**Hem.**	**Coordinate**			***T***	***Z***
				x	*y*	*z*		
Large > normal voice (go trials)	64	Supplementary motor area	L	-2	-4	70	7.63	5.33
	37	Inferior frontal gyrus, pars opercularis	L	-50	10	16	7.22	5.17
	73	Anterior cingulate cortex	L	-6	26	36	6.99	5.07
		Supplementary motor area	L	-6	18	42	6.57	4.88
	8	Anterior Insula	L	-36	24	0	6.38	4.79
Hostile > normal voice (go trials)	296	Supplementary motor area	R	6	-2	70	8.13	5.53
			R	4	6	68	7.44	5.26
		Superior frontal gyrus	L	-12	2	72	7.49	5.28
	130	Middle cingulate cortex	R	2	16	30	7.42	5.25
	78	Anterior Insula	R	42	12	0	7.92	5.45
	9	Anterior Insula	L	-32	10	-18	6.97	5.06
	5	Anterior Insula	L	-42	6	2	6.44	4.82
	47	Retrosplenial cortex		0	-32	6	6.68	4.93
	29	Corpus callosum	L	-4	-4	24	8.30	5.59
			R	4	-4	26	7.03	5.09
	10	Supramarginal Gyrus	R	62	-28	28	6.73	4.95

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. p<.05 FEW corrected, minimal cluster size: 5 voxels.

2.3 Study 3: What makes a vocal chameleon? Individual differences in social vocal control are reflected in activation of vocal motor and social processing regions.[3]

[3] Guldner, S., Nees, F., Flor, H., McGettigan, C. (2020). *What makes a vocal chameleon? Individual differences in social vocal control are reflected in activation of vocal motor and social processing regions.* Manuscript in preparation.

What makes a vocal chameleon? Individual differences in social vocal control are reflected in activation of vocal motor and social processing regions.

Stella Guldner[1,4*], Frauke Nees[1,2] , Herta Flor[1] and Carolyn McGettigan[3]

[1] Institute of Cognitive and Clinical Neuroscience, Central Institute of Mental Health, Medical Faculty Mannheim, Heidelberg University, Mannheim, Germany

[2] Institute of Medical Psychology and Medical Sociology, University Medical Center Schleswig Holstein, Kiel University, Kiel, Germany

[3] Department of Speech, Hearing and Phonetic Sciences, University College London, London, UK

[4] Graduate School of Social and Economic Science, University of Mannheim, Mannheim, Germany

Corresponding author: Stella Guldner, Central Institute of Mental Health, J5, 63159 Mannheim, email: stella.guldner@zi-mannheim.de, phone: +49 621 1703 6314

Abstract

In our everyday life we coordinate numerous social interactions using our voice. Such social vocal control relies on an interplay of functional processing in networks involved in vocomotor control and social interaction. Individual differences in social reactivity might support the efficacy with which social information is encoded in the voice. Moreover, the strategic modulation of the voice to express beneficial social traits (such as likeability) might contribute to the aptitude with which people with exploitative social styles engage in social interactions. Yet we understand little about the psychological and neurophysiological mechanisms that contribute to successful vocal modulation. We used functional magnetic resonance imaging (fMRI) to investigate the relationship of cognitive and affective empathy, psychopathic and Machiavellistic traits to social vocal control and its neural correlates.

Twenty-four right-handed, native British English speakers (twenty females) modulated their voice to communicate social traits (sounding likeable, hostile, and intelligent) while undergoing a rapid-sparse fMRI protocol. Performance in social vocal control was operationalized as the specificity with which speakers evoked trait percepts in an independent group of naïve listeners. Speakers' empathy levels, as well as psychopathic and Machiavellistic traits, were assessed using self-report questionnaires.

Better performance in social vocal control was related to increased functional activation in brain regions associated with socio-cognitive processing and vocal control (TPJ, SMG, premotor cortex). Machiavellistic traits were positively associated with portraying likeability in the voice, but this was not reflected in the neural level. Exploratory analyses showed that, during likeable voice production, lower cognitive and affective empathy as well as higher psychopathic traits were associated with higher activation levels in brain regions involved in social processing (orbital IFG, mPFC, STS TPJ, precuneus). These findings highlight the psychological and neural mechanisms involved in strategic social voice modulation, suggesting differential processing to achieve voice modulations in a combined network of vocal control and social processing streams.

Keywords: Vocal control, Social interactions, Empathy, fMRI, Machiavelliansim

Highlights:

- Ability to express specific social traits in voices is associated with activation in brain regions involved in vocal motor and social processing
- Social reactivity indices in speakers is not associated with performance in volitionally expressing social information in the voice
- Self-reported levels of Machiavellianism are associated with better performance in volitional vocal expressions of likeability
- Producing a likeable voice differentially engages brain areas involved in social processing depending on the speaker's self-reported cognitive and affective empathy levels.

What makes a vocal chameleon? Individual differences in social vocal control are reflected in activation of vocal motor and social processing regions

Vocal behaviour is a predominantly social signal and the primary channel for social communication. Listeners form impressions about social attributes and personality traits from speakers' voices rapidly (McAleer, Todorov, & Belin, 2014) and reliably (Mahrholz, Belin, & McAleer, 2018) along stereotypical dimensions of affiliation and competence (Fiske, Cuddy, Glick, & Xu, 2002; McAleer et al., 2014). These are often generalized to other personality characteristics (Hughes & Miller, 2016; Montepare & Zebrowitz-McArthur, 1987; Zuckerman & Driver, 1989), and have far-reaching social consequences (O'Connor & Barclay, 2017; Pavela Banai, Banai, & Bovan, 2017; Schroeder & Epley, 2015; Tigue, Borak, O'Connor, Schandl, & Feinberg, 2012; Torre, Goslin, White, & Zanatto, 2018). Thus, receiving favourable trait judgments is an important contributor to successful social interactions and achieving social goals. One route of doing this is through the control of nonverbal behaviours such as vocal modulations (DePaulo, 1992). Previous work has shown that speakers can volitionally amplify trait judgments in their voice, for instance modulating their voice to sound more dominant, intelligent or attractive (Hughes, Mogilski, & Harrison, 2014). Such social vocal control might therefore be important for speakers to manage their impression on others and engage in meaningful social interactions. However, it is yet unclear what psychological and neurophysiological characteristics determine a speaker's ability to voluntarily express social information in their voice, and in turn manage the perception of their voice by listeners.

Successful social communication has been suggested to rely on the ability to attribute mental states to others (Sperber, 2000), suggesting that the ability to understand and represent a listener's feelings, intentions and thoughts, can inform a speaker to behave in a way that addresses these feelings and thoughts adequately. This process is commonly referred to as cognitive empathy, thought to be under voluntary control (Bird & Viding, 2014) and closely related to theory of mind (ToM; Blair, 2005), whereas the vicarious experience of another's emotions is referred to as affective empathy (Reniers, Corcoran, Drake, Shryane, & Völlm, 2011). In fact, empathy has been shown to predict the accuracy with which authenticity is recognized in nonverbal vocalizations (Neves, Cordeiro, Scott, Castro, & Lima, 2018) as well as the perception and understanding of irony (Amenta, Noël, Verbanck, & Campanella, 2013; Jacob, Kreifelts, Nizielski, Schütz, & Wildgruber, 2016), and emotions (Aziz-Zadeh, Sheng, & Gheytanchi, 2010). In populations where empathy is selectively impaired, such as in autism spectrum disorders (ASD) or psychopathy, vocal behaviour is characteristically changed and often described as sounding odd or emotionally flat (Fusaroli, Lambrechts, Bang, Bowler, & Gaigg, 2016; Louth, Williamson, Alpert, Pouget, & Hare, 1998; Shriberg et al., 2001). Although spontaneous abnormal vocal behaviour is observed both in ASD and psychopathy, this might be different for volitional vocal behaviour. Psychopaths, for instance, seem to be more effective and persuasive in volitionally expressing emotions through facial expression compared to control participants (Book et al., 2015). This might be due to differential impairments in empathy in these two populations: whereas ASD is associated with specific deficits in cognitive empathy, psychopathy is associated with deficits in affective empathy (Lockwood, Bird, Bridge, & Viding, 2013). In fact, the ability to coolly apprehend others' feelings and thoughts (cognitive empathy) without sharing the emotional experience (affective empathy), might in a malevolent context contribute to the aptness with which people with high psychopathic tendencies manipulate and deceive others (Austin, Farrelly, Black, & Moore, 2007).

Another personality trait that might be associated with good nonverbal communication skills might be Machiavellianism (MACH). MACH traits are defined by a highly strategic interpersonal style, which is mostly self-serving and manipulative, whereas psychopathic traits are associated with a more manipulative interpersonal style, which is often callous, unemotional and impulsive (Fehr, Samson, & Paulhus, 1992; Jones & Paulhus, 2009). While some work shows a positive relation between cognitive empathy components (e.g. perspective-taking) and MACH (Szabó & Bereczkei, 2017), other studies show no or negative associations with both cognitive and affective empathy skills in subclinical samples (Jonason & Krause, 2013; Nagler, Reiter, Furtner, & Rauthmann, 2014; Wai & Tiliopoulos, 2012). Nevertheless, in dyadic interactions, higher levels of Machiavellianism and self-monitoring are associated with stronger linguistic alignment, specifically if it serves the interest of the speaker (Muir, Joinson, Cotterill, & Dewdney, 2016). Therefore, it remains to be elucidated whether volitional social vocal behaviour might be a route through which psychopaths and Machiavellists manage others' impressions.

The neural mechanisms supporting the ability to volitionally modulate the voice to express social information remain elusive. The voice is a voluntary and highly dynamic behaviour, which results from an intricate interplay of fine-tuned motor commands (Guenther, 2006). These are controlled on the neural level through a network of regions involved in vocomotor control, often referred to as the vocomotor network (VMN), including left inferior frontal gyrus (IFG), anterior cingulate cortex (ACC), supplementary motor cortex

(SMA), supramarginal gyrus (SMG), superior termporal gyrus (STG), insula, basal ganglia (BG) and cerebellum (reviewed by Pisanski, Cartei, McGettigan, Raine, & Reby, 2016; Simonyan & Horwitz, 2011). Control over the vocal apparatus (larynx, vocal tract, articulators), through the VMN allows for filtering and shaping of the subglottal airflow into specific frequencies, vocal quality and loudness (Fant, 1960). Accordingly, vocal behaviour is limited to certain individual anatomical boundaries. Nevertheless, the voice is also strikingly dynamic, with substantial within-speaker variation (McGettigan, 2015) that allows to adjust vocal behaviour to the social environment (e.g. Burnham, Kitamura, & Vollmer-Conna, 2002; Farley, Hughes, & LaFayette, 2013; Kemper, Finter-Urczyk, Ferrell, Harden, & Billington, 1998).

Previously, we found support for the hypothesis that neural areas involved in social cognition might support the volitional expression of social traits in the voice (Guldner, Nees, & McGettigan, 2020). Empathy, and in particular mentalizing about others' emotions, intention, beliefs and thoughts (Frith & Frith, 2003) reliably engages the superior temporal sulci, temporo-parietal junction (TPJ), medial prefrontal cortex (mPFC) as well as precuneus (Bzdok et al., 2012; Molenberghs, Johnson, Henry, & Mattingley, 2016; Schurz, Radua, Aichhorn, Richlan, & Perner, 2014). These regions are often referred to as the social brain network (SBN). SBN activation is related to both metalizing about others and the self (Lombardo et al., 2010; Nicolle et al., 2012) and particularly ventral mPFC and precuneus are implicated in trait processing (Ma et al., 2014; Van Overwalle, Ma, & Baetens, 2016) in relation to the social trait space dimensions of competence and affiliation (Cuddy, Fiske, & Glick, 2008). Ventral mPFC might also support predicting future social outcomes based on social experiences (Krueger, Barbey, & Grafman, 2009; Thornton, Weaverdyck, & Tamir, 2019). In relation to vocal behaviour, SBN activation supports explicit (McGettigan et al., 2013) as well as covert (Brown, Cockett, & Yuan, 2019) identity expression through vocal modulations. Moreover, activation in superior temporal sulci during affective voice production is modulated by the social context (i.e. the presence of others; Klasen et al., 2018). SBN activation might therefore support the ability to express social trait information in the voice, but individual differences in social vocal control ability in relation to social reactivity have not been tested previously. An alternative hypothesis would be that engagement of the VMN predicts social vocal control performance. This hypothesis is supported by evidence that the ability to volitionally modulate the voice to express sad or happy emotions is associated with increased functional activation of right IFG (triangular part), right middle frontal gyrus and left superior frontal gyrus that might be related to motor planning for prosody production (Aziz-Zadeh et al., 2010).

One study has directly investigated the correlates of affective empathy and psychopathy traits with functional processing during the perception of affective prosody (Aziz-Zadeh et al., 2010). The authors report activation in left IFG to be significantly associated both with affective empathy traits and affective subcomponents of psychopathic traits. Psychopathic traits were additionally negatively associated with activation in other SBN areas, including TPJ, STS and the bilateral anterior insulae. However, the authors tested only affective subcomponents of empathy and psychopathic personality traits, and did not test whether the association between affective empathy and affective prosody perception also extended to prosody production, with prosody perception and production having been shown to involve common processing areas in IFG (Aziz-Zadeh et al., 2010).

In summary, most previous studies have used either combined measures of empathy (Amenta et al., 2013; Jacob et al., 2016), focussed on affective empathy components (Aziz-Zadeh et al., 2010; Neves et al., 2018) or - most importantly - studied individual differences in empathy in relation to perception but not voice production. Moreover, these studies mostly refer to emotional voice changes, but not social voice changes. It is therefore unclear how cognitive or affective empathy components in relation to socially opportunistic behaviour might contribute to the ability to volitionally express social information through the voice, and which neurophysiological mechanisms might support this ability.

Here, we investigate individual differences in the ability to convey social traits through the voice in relation to individual differences in social reactivity, including dispositional empathy (cognitive and affective empathy), and socially-opportunistic personality traits (Machiavelliansim and psychopathy). First, we hypothesized that performance in social vocal control would be positively associated with empathy (particularly cognitive empathy) and socially-opportunistic personality traits. Given, that these measures are interrelated, we tested this in the framework of multiple regression models, to determine each individual and unique contribution. Further, we probed the underlying neurophysiological networks associated with individual differences in social vocal control, hypothesizing a positive association between activation areas associated with domain-general social processing (mPFC, STS/TPJ, precuneus) during social voice modulation. Lastly, we hypothesized that activation in these SBN regions as well as their functional connectivity with left IFG (in the VMN) will be positively associated with social reactivity indices during social voice modulation. Given the exploratory nature of our study, we explored brain-behaviour associations based on first establishing whether there was a behavioural association between the personality traits of the speaker and social vocal control ability. For the interest of the reader, we report associations with all individual differences measures.

Methods

Speakers. Twenty-four right-handed, native British English speakers (M_{age} = 21.04 SD=3.26, 3 male) participated in this experiment. All speakers had normal or corrected-to-normal vision and reported no history of hearing, language, neurological or psychiatric disorders and were recruited from the participant pool at the Department of Psychology at Royal Holloway, University of London and received 30 GBP as reimbursement. All speakers provided their full informed and written consent prior to participation according to the Declaration of Helsinki (1991). This study was approved by the research ethics committee of Royal Holloway, University of London (587-2017-10-24-14-50-UXJT010).

Assessment of Empathy, Machiavellianism and Psychopathy. To measure individual differences in trait empathy, speakers completed the Questionnaire of Cognitive and Affective Empathy (QCAE; Reniers, Corcoran, Drake, Shryane, & Völlm, 2011). The QCAE is a 31-item questionnaire measuring cognitive empathy in two subscales including perspective taking and online simulation, and affective empathy on three subscales of emotion contagion, proximal responsivity and peripheral responsivity. Speakers choose their level of agreement with each item from 1 (*strongly disagree)* to 4 (*strongly agree*). To assess traits of psychopathy and Machiavelliansim, we used the Short Dark Triad (SD3; Jones & Paulhus, 2014), which is a 27-item self-report questionnaire. Here, we focused on dispositional psychopathy and Machiavellianism, as these constructs are highly associated with social manipulative behaviours. Responses are given in agreement with each item on 5-point Likert-scale with anchors 1 (*strongly disagree*) to 5 (*strongly agree*).

Social vocal control task. The main experimental task consisted of a social vocal control task (Guldner et al., 2020) in which speakers were asked to express social and non-social traits in the voice. Social traits included vocally expressed intelligence, likeability and hostility. Modulating the voice to express a large body size, as well as speaking in non-modulated neutral voice, were implemented as control conditions (for instructions on trait expressions see Guldner et al., 2020). A body-size modulation was chosen as a non-social control condition because it demands comparable vocal tract manipulations (e.g. larynx lowering, changes in vocal fold tension; Pisanski et al., 2016). Exemplars consisted of four two-syllable, five-letter pseudowords with a C-V-C-V-C (C=consonant, V= vowel) phonotactic structure (*belam*, *lagod*, *minad*, and *namil*; Frühholz et al., 2015).

Design and procedure. Speakers filled out the self-report questionnaires online prior to the fMRI scanning session. One the scanning day, speakers first received a brief training on the vocal control task before completing the task in the MRI scanner over the course of 4 runs. Each run consisted of a randomized order of the 5 vocal modulation conditions (normal / large / hostile / likeable / intelligent) paired with one of the 4 exemplars, of which each combination appeared during 3 Go trials and 3 No-Go trials over the course of one run. Go and No-Go trials were presented in randomized order, with the restriction of a maximum of three No-Go trials in a sequence to avoid a decrease in attention. In total, each run included 150 trials of which 30 were rest trials.

Both Go and No-Go trials started with a two-second presentation of the target trait and a fixation cross, allowing speakers to prepare expressing the targeted trait in their voice. However, only during Go trials, the fixation cross was then substituted with an exemplar for 1.5 seconds. During this 1.5 sec silent gap speakers were asked to vocalize the exemplar while expressing the target trait and utterances were recorded using an in-scanner MR-compatible microphone (Opto-acoustics, FOMRI-III). During No-Go trials, on the other hand, the fixation cross remained on the screen for the duration of the silent gap and no exemplar was presented. Go and No-Go trials therefore allowed filtering out neural activation specifically related to ongoing voice production (Go) while controlling for any other task-related cognitive processing effects (present in Go and No-Go). Visual cues were projected onto a screen at the back of the scanner bore and viewed via a mirror on the headcoil. All stimuli were presented using the Psychophysics toolbox (Kleiner, Brainard & Pelli, 2007) in Matlab (2014a, the Mathworks, Natick, MA). The total scanning time was approximately 50 minutes.

Based on their voice recording acquired in the scanning session, speakers were re-invited approximately 1 week later to rate how strongly they had expressed the social trait on a 7-point Likert scale ranging from not at all (=1) to very much (=7). To do this, they were presented with their voice recordings blocked by the expressed target trait in a soundproof booth via Sennheiser Headphones HD306, using the Psychophysics toolbox (Kleiner, Brainard & Pelli, 2007) in Matlab (2014a, the Mathwork, Natick, MA). Based on these ratings, we selected one recording for each modulation condition and for each speaker that

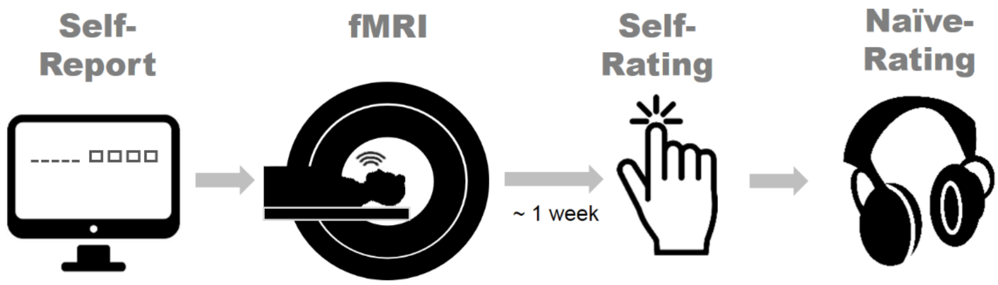


Figure 5. Experimental Procedure.

had received the speaker's own maximal rating. This was done to ensure that independent evaluations of performance would be based on sounds where the original speakers had felt confident in their accurate expression of the target trait. In cases where multiple recordings had the same maximal rating, we randomly selected one of these recordings. The experimental procedure is illustrated in Figure 1.

Naïve ratings. For each speaker, the highest rated recording for each trait category was intensity normalized across speakers (Guldner et al., 2020) and then presented to 24 naïve listeners. All raters (M_{age}=19.92, SD=1.47, 4 male) were recruited at the Department of Psychology at Royal Holloway, University of London, gave their informed consent prior to participation and received monetary compensation for their participation of 5 GBP.

To reduce the experimental duration, each rater heard the recordings of a subset of 10 speakers, while ensuring that each speaker was heard by at least 10 different raters. For each speaker we included one recording of each vocal modulation condition (trait) and one recording of their neutral voice. Each rater heard and rated each recording of each speaker on all trait scales in separate blocks. Each block consisted of one social trait rating for all recordings in randomized order (e.g. one block for rating all recordings on likeability; additional ratings of arousal and valence are reported elsewhere (Guldner et al., 2020). Raters indicated their responses on 7-point Likert scales, with anchors at 1 (*not at all*) to 7 (*very*) to rate how strongly a voice expressed a given trait. Block order was randomized across raters. Stimuli were presented using the Psychophysics Toolbox (Kleiner, Brainard & Pelli, 2007) in Matlab (2014a, the Mathwork, Natick, MA) on Sennheiser Headphones (Sennheiser U.K. Ltd, Marlow, UK) in a soundproof booth.

Individual differences in social vocal control. The ratings obtained from naïve raters were analysed in R (http://www.R-project.org/). To assess the success of the voice manipulations to express social traits, we calculated the average change in naïve ratings for modulated voices relative to the normal voice samples, for each speaker and each intended trait (i.e. comparing "intelligent" ratings for the normal and "intelligent" trials), thus allowing to compute a performance index for each trait for each speaker (henceforth Δ - [trait]).

Next, we computed representational similarity matrices (RSM) based on the pairwise Pearson correlation coefficients of naïve ratings between pairs of social traits (likeability, hostility and intelligence). These matrices permit to characterise the similarity in mental representations between different stimulus categories (Kriegeskorte, Mur, & Bandettini, 2008); see e.g. Kuhn, Wydell, Lavan, McGettigan, & Garrido, 2017; Sauter, Eisner, Calder, & Scott, 2010 for similar approaches). In this study, we made use of this analysis to explore how specific voice modulations were expressed in respect to the trait percept they induced (reflected in the naïve ratings). Each cell of the matrix contained the pairwise correlations of mean Δ-ratings for the respective two trait categories. From these matrices, we computed a general performance parameter to capture specificity of social voice modulations for each speaker. This parameter was estimated by calculating the Euclidean distance of each speaker's RSM to a theoretical matrix with maximized discrimination between trait ratings (see Figure 2). Based on the social space dimensions shown elsewhere (Guldner et al., 2020; McAleer et al., 2014), we assumed that there would be no significant correlations ($r=0$) of intelligence ratings with either likeable or hostile voice modulation, and that the ratings for hostile and likeable voices would be anti-correlated ($r=-1$).

Effects of social reactivity on performance in social vocal control. To investigate the effect of social reactivity indices (self-reported cognitive and affective empathy, psychopathic and MACH traits) on vocal modulation performance, we calculated a multiple regression model to test whether social reactivity indices are predictive of social vocal control ability (Euclidean distance measure). The model included the social reactivity indices as regressors, and age and gender as covariates of no interest. On the trait level, we calculated partial Pearson correlation coefficients pairwise for each spoken trait performance (Δ - [trait]; e.g. Δ-intelligence) and each social reactivity index, controlling for age and gender. We corrected for multiple comparisons using the False Discovery Rate (FDR; Benjamini & Hochberg, 1995).

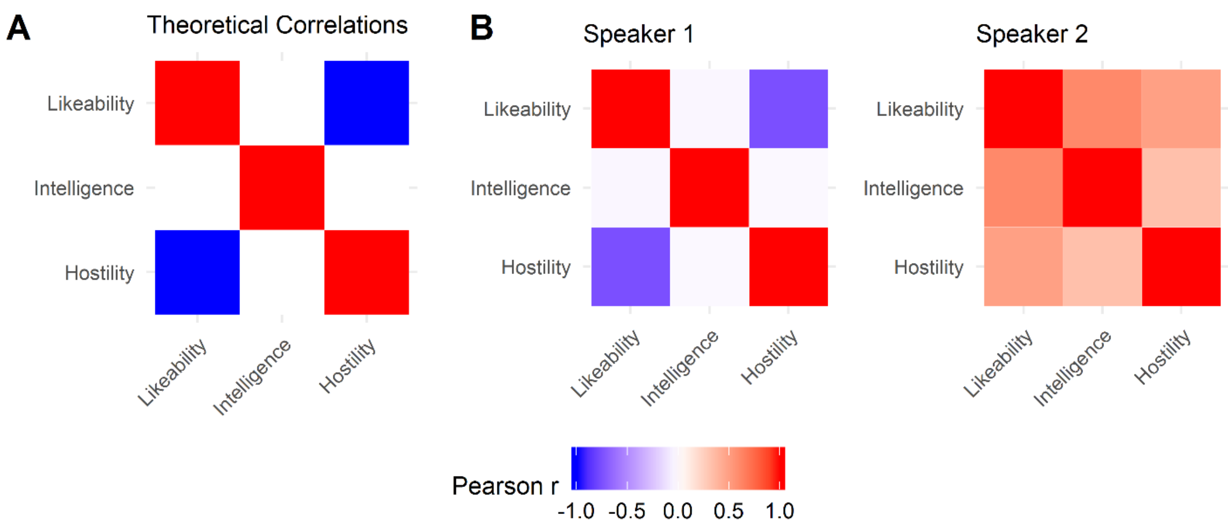


Figure 6. A. Theoretical Representational Similarity Matrix (RSM), showing expected correlations between evoked trait ratings. **B.** Exemplary Individual RSMs for two speakers, with high (left) and low (right) specificity between evoked trait ratings.

Neural correlates of performance in social vocal control. Functional brain images were preprocessed and analysed as described previously in detail (Guldner et al., 2020; see supplementary materials S1). To determine brain regions associated with performance in social vocal control, we ran a multiple regression model on the group level for the contrast "Social Go > Rest" with performance in social vocal control (Euclidean distance) as a covariate of interest.

Effect of social reactivity on task-based functional activation. To explore whether differences in behavioural performance related to social reactivity were associated with differences in neural processing during social voice modulation, we calculated a multiple regression model based on the behavioral association on the contrast "[trait] Go > Rest" and the social reactivity indices as covariates of interest. All models included a constant intercept, and age and gender as covariates of no interest.

Effect of social reactivity on task-based functional connectivity. Lastly, we explored the effect of social reactivity indices on task-based functional connectivity with a core vocomotor control area in left IFG [-52 12 24], during Go versus No-Go trials. Psychophysiological interactions (PPI) were computed in SPM12 following the procedure described previously

(Guldner et al., 2020; see supplementary materials S2). On the group level, we computed a multiple regression model including the social reactivity indices as regressors as well as a constant intercept, age and gender as covariates of no interest. This approach allowed us to estimate the effect of social reactivity on functional connectivity during ongoing social vocal control to express social traits between left IFG and other brain areas while cancelling out other task-related effects unrelated to voice production (Go vs No-Go trials). Thus, we targeted the association between social reactivity indices and functional connectivity specifically involved in achieving ongoing social voice production.

For all imaging analysis, we used a significance threshold of uncorrected $p<.001$ on the voxel level, corrected for multiple comparisons on the cluster level with $p<.05$ using an individual cluster extent threshold for each contrast, which we determined using a Monte-Carlo simulation with 1000 iterations (Slotnick, Moo, Segal, & Hart, 2003). Resulting clusters were labelled based on the location of each peak activation using the in-built Neuromorphometrics and the automated anatomical labeling (AAL) atlas in SPM12.

Results

Individual differences in social vocal control

Descriptive statistics of the behavioural measures can be found in Table 1. Performance social vocal control (Euclidean distance from theoretical RSM) ranged from 0.25 to 1.64, with M=0.74, SD=.034, where smaller values denote higher specificity in vocally evoked trait ratings in naïve listeners (see supplementary materials S4). On the trait level, speakers evoked a mean increase of 1.30 points on the Likert-scale with their likeable voice modulations relative to their neutral voice (Δ-Likability: M=1.30, SD=0.66, $t(23)=5.64$, $p<.001$), an increase of 2.07 for hostile voice (Δ-Hostility: M=2.07, SD=1.41, $t(23)=6.68$, $p<.001$), and an increase of 0.53 for intelligent voice modulations (Δ-Intelligence: M=0.53, SD=0.87, $t(23)=1.87$, $p<.05$; results of multivariate trait rating comparisons have been reported previously (Guldner et al., 2020).

Table 1. Descriptive statistics of social reactivity and performance in social vocal control and partial Pearson correlation coefficients, controlling for speaker gender and age.

	M (SD)	α	1	2	3	4	5	6
1 Machiavellianism	25.29 (4.79)	0.74	-					
2 Psychopathy	16.92 (3.86)	0.61	**0.60***	-				
3 Cognitive Empathy	57.42 (8.32)	0.91	0.06	-0.30	-			
4 Affective Empathy	36.67 (5.15)	0.81	-0.27	-0.19	-0.00	-		
5 Δ-Hostility	2.07 (1.41)	-	0.08	0.10	-0.01	-0.07	-	
6 Δ-Intelligence	0.53 (0.87)	-	0.13	0.33	-0.32	-0.03	-0.10	-
7 Δ-Likeability	1.30 (0.66)	-	**0.66 ****	0.53[t]	0.20	0.01	-0.19	0.05
Euclidean Distance	0.74 (0.03)	-						

Note. * $p \leq .05$ ** $p \leq .01$ FDR corrected, [t] $p=.08$ FDR corrected; FDR= False discovery rate, M= Mean, SD= Standard deviation; α = Cronbach's alpha.

Effects of social reactivity on performance in social vocal control

The multiple regression model showed no significant association between social reactivity indices and general performance in social vocal control (Euclidean Distance; adj.R^2=0.0, $F(6,17)=0.91$, $p=.51$ all βs<-.31, all ts(17)<1.61, all ps>.13). We next explored associations between social reactivity and performance in social vocal control on the trait level and found a significant positive association between Δ-Likeability and Machiavellianism (r_p= .66, $p<.02$). Performance in all other traits was not significantly associated with psychopathic or Machiavellistic traits, nor were there significant correlations of performance with cognitive or affective empathy (all rs<.32, all ps>.08, see Table 1). To explore this result further, we employed an additional multiple regression model including Δ-Likeability as the outcome variable and all social reactivity indices as predictor variables, while controlling for age and gender. The model explained 46 % of the variance in Δ-Likeability (adj.R^2=.46, $F(6,17)=4.21$, $p<.01$) and revealed that levels of Machiavellianism predicted likeable voice performance independent of other social reactivity indices ($\beta=.54$, $t(17)=2.30$, $p<.05$).

Neural correlates of performance in social vocal control

The whole-brain multiple regression model showed a significant association between functional activation and general performance in social vocal control (Euclidean Distance) in 9 clusters (uncorrected $p<.001$, $k=61$) with peak voxels in left posterior Temporo-parietal junction (TPJ), right middle frontal gyrus (MFG), bilateral supramarginal gyrus (SMG), left middle temporal gyrus (MTG), left somatosensory cortex in postcentral gyrus (S1), cuneus, and bilateral caudate (see Table 2 and Figure 3).

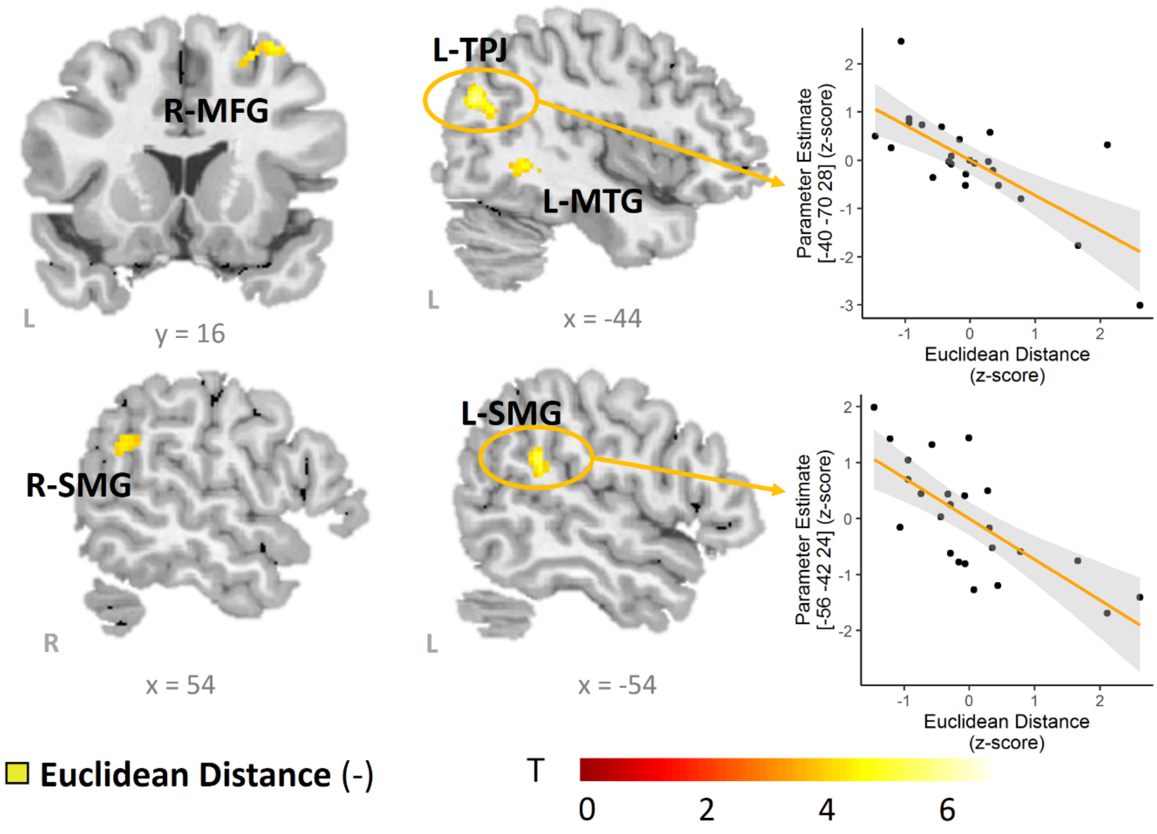


Figure 7. Correlates of Social Vocal Control Ability: Activation Maps. The multiple regression model revealed a significant negative association between the Euclidean distance measure and functional activation in response to Social Go trials > Rest in left posterior TPJ, MTG and right MFG, as well as bilateral SMG. Associations are illustrated based on the peak-voxel parameter estimates in left posterior TPJ and SMG. MFG=middle frontal gyrus, MTG=middle temporal gyrus, TPJ=temporo-parietal junction, L=left, R=right.

Effect of social reactivity on task-based functional activation

Given that Machiavellistic and psychopathic tendencies were associated with better performance in expressing likeability (Δ-Likeability), we explored possible underlying differences in functional processing in networks involved in social processing. We ran a multiple regression analysis on the contrast Likeable Go > Rest with the questionnaire indices as regressors, while controlling for gender and age. We found no significant clusters associated with Machiavellianism. However, given that these analysis were exploratory, we report associations with all social reactivity indices.

Table 2. Functional activations for the multiple regression whole-brain model of performance (Euclidean distance) on Social Go > Rest. All contrasts are negative correlations, the positive direction showed no significant voxels.

Contrast	*k*	Region	Hem.	x	y	z	*T*	*Z*
(-) Euclidean	338	Caudate	L	-18	-22	22	6.31	4.63
Distance on			L	-8	-26	16	4.77	3.85
Social Go > rest	295	Posterior Temporo-Parietal Junction	L	-40	-70	28	5.68	4.33
			L	-38	-62	20	5.13	4.05
			L	-28	-74	44	3.95	3.36
	221	Caudate	R	18	-10	22	4.59	3.75
			R	20	-26	18	4.19	3.51
			R	16	-30	10	4.05	3.42
	105	Postcentral Gyrus/S1	L	-36	-28	40	4.72	3.82
			L	-32	-26	48	4.17	3.49
	96	Middle Frontal Gyrus	R	32	14	60	4.37	3.62
			R	30	6	62	3.98	3.37
		Superior Frontal Gyrus	R	24	16	56	4.07	3.43
	84	Middle Temporal Gyrus	L	-42	-50	0	4.40	3.64
			L	-42	-40	-2	3.98	3.38
	69	Supramarginal Gyrus	R	50	-50	30	4.63	3.77
	68	Supramarginal Gyrus	L	-56	-42	24	4.91	3.93
	63	Cuneus	L	-14	-80	20	4.22	3.53

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right, (-) denotes negative direction. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. $p<.001$ uncorrected, minimal cluster size: 61 voxels.

Higher affective empathy was associated with increased activation during likeable voice production in the left posterior TPJ, ventral medial prefrontal cortex (mPFC) and precuneus (PrCu). Lower cognitive empathy was associated with increased activation in a cluster in the dorsal mPFC/anterior cingulate cortex (ACC), right STS, left parahippocampal gyrus (pHC), right orbital inferior frontal gyrus (IFG) and PrCu. Higher psychopathic trait scores were associated with decreased activation in an overlapping region of the orbital IFG, and a region in the temporal pole (uncorrected p<.001, k=60; see Table 3 and Figure 4).

Effect of social reactivity on task-based functional connectivity

Lastly, we explored whether personality traits were associated with patterns of functional connectivity between left-IFG and social processing areas. The multiple regression model revealed no significant effects of Machiavellianism, nor psychopathic, or empathy traits.

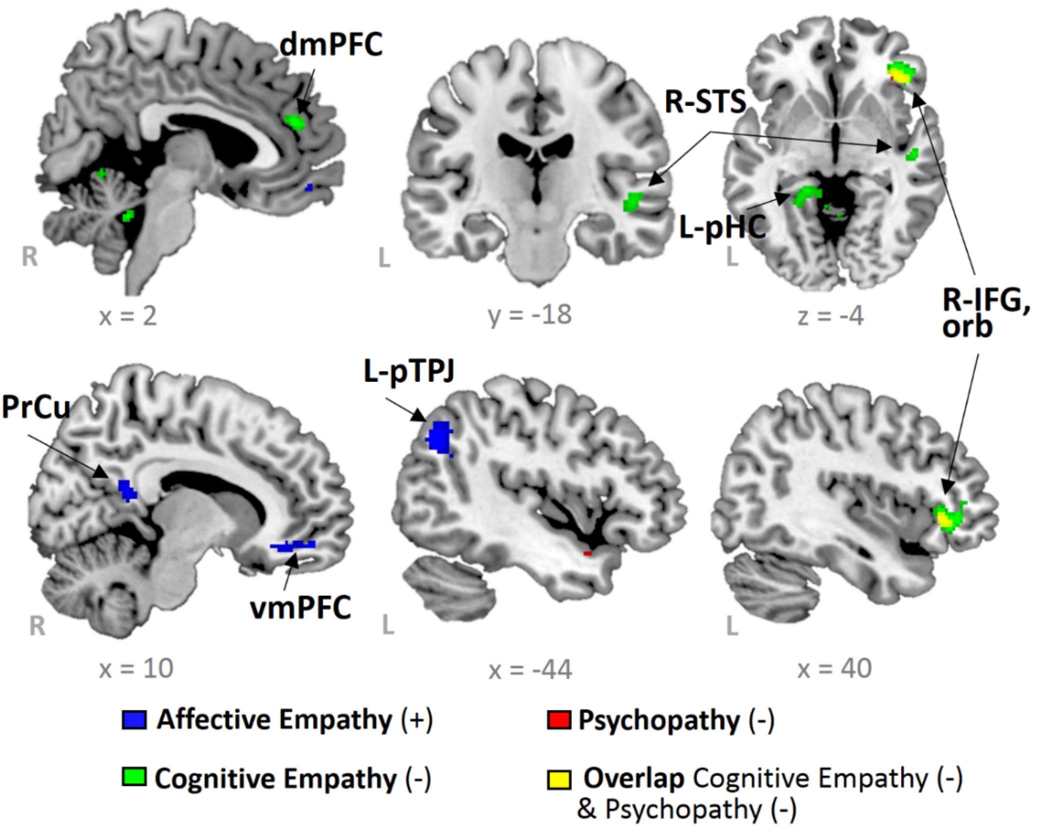


Figure 8. Activation maps: Effect of social reactivity on functional activations while speaking in a likeable voice. Negative associations between task-based functional activation and cognitive empathy (green) were found in right IFG, orbital part, right STS, left paraHC and dmPFC. The cluster in orbital parts of right IFG was also negatively associated with psychopathy (red; overlap shown in yellow). Affective empathy was positively associated with activation in left pTPJ, PrCu, vmPFC. IFG= inferior frontal gyrus, dmPFC=dorsomedial prefrontal cortex, pHC=parahippocampal gyrus, PrCu=precuneus, pTPJ=posterior temporo-parietal junction, STS=superior temporal sulcus, vmPFC=ventromedial prefrontal cortex, L=left, R=right.

Discussion

In this study, we aimed to explore individual differences in the ability to volitionally express social information in the voice. We tested the influence of dispositional empathy and socially-opportunistic personality traits on social vocal control ability, as well as associated neural activation differences during voluntary socially motivated voice changes. We found that general performance in social vocal control was not associated with differences in self-reported cognitive or affective empathy, psychopathy and Machiavellianism. On the neural level, individual differences in performance in social vocal control were associated with increased functional activity during social voice changes in areas associated with social processing and vocal motor control. Performance in expressing favourable social traits (sounding likeable) was uniquely and positively associated with MACH traits. However, we

Table 3. Functional activations for the multiple regression whole-brain model of personality traits on Likeable Go > Rest.

Contrast	*k*	Region	Hem.	x	y	z	*T*	*Z*
(+) Machiavellianism	-							
(-) Machiavellianism	-							
(+) Psychopathy	-							
(-) Psychopathy	85	Temporal Pole	L	-32	8	-22	4.49	3.59
			L	-42	10	-22	4.43	3.56
	71	Inferior Frontal Gyrus, orbital part	R	42	32	-4	6.20	4.42
(+) Cognitive Empathy	91	Inferior Parietal lobe	L	-36	-36	28	5.52	4.12
(-) Cognitive Empathy	235	Inferior Frontal Gyrus, orbital part	R	42	32	-4	9.11	5.42
			R	42	44	6	4.32	3.50
	126	Parahippocampal Gyrus	L	-24	-46	-6	4.89	3.81
		Cerebellum	R	4	-58	-6	4.08	3.36
			L	-10	-44	-6	4.00	3.31
	89	Inferior Frontal Gyrus / Anterior	R	32	14	-12	4.90	3.82
		Insula	R	32	4	-10	3.98	3.30
	88	Dorsal Medial Prefrontal Cortex / Anterior Cingulate Cortex	R	4	46	20	5.43	4.08
	74	Cerebellum	R	4	-44	-28	5.12	3.93
			L	-4	-48	-32	4.29	3.48
	68	Precuneus	L	-12	-56	22	6.32	4.47
	65	Superior Temporal Sulcus	R	50	-16	-4	4.89	3.81
			R	44	-24	2	3.74	3.15
(+) Affective Empathy	236	Posterior Temporo-Parietal Junction	L	-42	-66	36	7.24	4.83
	93	Precuneus	R	12	-46	18	5.38	4.05
			R	16	-48	28	4.51	3.60
	74	Ventral Medial Prefrontal Cortex	R	8	50	-14	5.73	4.22
			R	12	38	-12	4.53	3.62
(-) Affective Empathy	-							

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right, (-) denotes negative direction, (+) denotes positive direction. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. $p<.001$ uncorrected, minimal cluster size: 60 voxels.

found no significant association between MACH traits and functional processing during likeable voice changes to explain better performance in likeable voice modulation. Exploratory analyses instead revealed spatially separable regions associated with cognitive and affective empathy traits during likeable voice modulation, possibly suggesting differential task strategies. Lastly, we found no significant effect of Machiavellianism, psychopathy or empathy components on functional connectivity with the core region of VMN in left IFG.

Measuring individual differences in social vocal control. We introduced a novel operationalization of social vocal control performance, in the form of RSMs, which reflect the differentiability of expressed traits based on the speakers' vocal modulations. Similar approaches have been used to examine the representation of emotion categories during voice processing (e.g. Kuhn et al., 2017; Sauter et al., 2010). Here we applied the reverse approach, and showed that untrained speakers can influence discriminability between *social* trait categories. This approach allowed us to compare an optimal modulation performance (maximal evoked discrimination between voice modulations for each trait) with individual speaker similarity patterns. We thus operationalized general social vocal control ability as the Euclidean distance between each individual speaker-evoked RSM based on the naïve listener ratings and a theoretical RSM with maximal differentiation between the social traits in a two-dimensional trait space (Fiske, Cuddy, & Glick, 2007). Thus, the Euclidean distance measure primarily indicates the specificity of trait percepts evoked by the speakers' voice modulations. We used a novel methodological approach and suggest that multivariate listener ratings can be a viable tool to study the specificity and sensitivity of voice modulation. Social interactions are in themselves highly complex and include a certain level of ambiguity (for instance, reflected in differential processing of ambiguous social voice information (Hellbernd & Sammler, 2018). This underlines the importance to include both specificity and sensitivity measures of social vocal control in future investigations, to allow us to measure both how clearly discernable the social information is to listeners (specificity), but also how strongly speakers can express that information in their voice (sensitivity).

Neural networks supporting performance in social vocal control. One aim of this study was to investigate functional activation associated with individual differences in social vocal control. We found that better performance in social vocal control (evoking specific trait percepts in independent listeners) was supported by an increase in activation in left posterior TPJ, bilateral SMG, right premotor cortex, left premotor cortex, somatosensory cortex and cuneus, suggesting the involvement of both vocomotor-related and social processing to support specific social trait expressions through the voice. This is partly in line with findings reported by Aziz-Zadeh et al. (2010), who reported an association between emotional prosody production performance with motor and affect related processing areas such as right IFG, MFG and left SFG. The partly diverging findings might be related to differences in the operationalization of performance: while Aziz-Zadeh et al. (2020) determined performance based on the expressed magnitude of affect in the voice rated by a single rater, we measured the specificity of volitional vocal behaviour in a multivariate approach using multiple independent raters.

Thus, while we corroborate the involvement of motor related processing streams for social voice changes in the right MFG, we offer novel evidence suggesting additional engagement of the left posterior TPJ and bilateral SMG in individual differences in social vocal control. SMG supports both auditory and somatosensory feedback integration during speech (Golfinopoulos et al., 2011; Oberhuber et al., 2016), while TPJ is a region reliably involved in domain-general cognitive and affective social processing (Bzdok et al., 2013; Schurz et al., 2014; Van Overwalle, 2009) for instance during perspective-taking (Wang, Callaghan, Gooding-Williams, Mcallister, & Kessler, 2016). Specifically posterior regions of TPJ are associated with socio-cognitive processing and social trait judgments (Schurz et al., 2014) and left pTPJ is suggested to support top-down control of endogenous episodic memory retrieval (Ciaramelli, Grady, Levine, Ween, & Moscovitch, 2010). Moreover, left pTPJ is engaged during life conversation (Jasmin et al., 2019) and volitional vocal impersonations (McGettigan et al., 2013). Globally, TPJ activation might support current and future mental state attribution (Cabeza, Ciaramelli, & Moscovitch, 2012) and construing a social context for behaviour (Carter & Huettel, 2013) in correspondence to immediate intentions and goals (Saxe & Wexler, 2005). Structurally, pTPJ is closely connected to regions in precuneus and vmPFC, which harbour social trait representations (Van Overwalle et al., 2016) and social scripts (Krueger et al., 2009). Our data suggest that, while precuneus and mPFC might support social vocal control (Guldner et al., 2020), the engagement of left pTPJ is predictive of the efficacy with which social traits are expressed in the voice. In our task, pTPJ activation might be related to creating an internal social context to allow encoding specific traits in the voice to influence the listener’s impression of the speaker. In fact, left TPJ has been found to support regulating another’s emotions through vocalizations (Hallam et al., 2014). However, the Euclidean distance measure reflects speaker evoked perceptual trait differentiation, and TPJ activation might thus also support social feedback processing during ongoing social voice production. In favour of this, left TPJ is essentially involved in monitoring and differentiation self from other produced speech (Mondino, Poulet, Suaud-Chagny, & Brunelin, 2016) and specifically engaged during intention inference from voices (Hellbernd & Sammler, 2018). Although future studies are needed to determine the exact contribution of these regions to social vocal control, left pTPJ might work in accordance with VMN areas to support encoding social information in the voice (McGettigan et al., 2013).

In summary, the network of regions reported here might reflect the speakers' ability to monitor and encode social trait expression during ongoing social voice production (TPJ), through somatosensory and auditory feedback processing (SMG, S1) and vocal motor planning (MFG). We thus offer first evidence that social processing areas, in particular TPJ, support the ability with which speakers control listeners' perception of their voice.

Association between social reactivity and social vocal control ability. Our behavioural data showed no significant association between neither MACH and psychopathic traits, nor trait cognitive and affective empathy with general performance in social vocal control. This might be due to several reasons. First, most previous work focussed on perceptual processing in socio-emotional vocal expression (e.g. Amenta et al., 2013; Aziz-Zadeh et al., 2010; Jacob et al., 2016; Neves et al., 2018), or manipulated socio-emotional vocal expression only indirectly (for instance in clinical populations with impairments in social reactivity: Louth et al., 1998; Shriberg et al., 2001). Therefore these studies might target more reactive vocal behaviour than in our experimental design. Given the reviewed literature, both cognitive and affective empathy might thus be related to spontaneous social vocal expression, but not necessarily volitional expression. Secondly, socio-cognitive processing might be more important in live interaction (Redcay & Schilbach, 2019), that is, when speakers have a listener and have to flexibly adjust their vocal behaviour accordingly. Since our in-scanner task required speaking in isolation, we might therefore not have evoked a comparable social processing load. However, in live interactions during which linguistic style accommodation was tested, there were no significant effects of empathy on the likelihood of accommodation (Muir et al., 2016), although nonverbal vocal characteristics were not assessed and cognitive and affective empathy were not distinguished. We suggest that future studies should examine both spontaneous and volitional vocal behaviour in live interactions to elucidate the relation between social reactivity (both cognitive and affective empathy), perceptual acuity and vocal modulation.

The social vocal chameleon. We also explored correlations between social reactivity indices and vocal control ability in each of the traits and found that the magnitude of the likeable trait percept evoked by the voice modulations was positively associated with MACH traits. Post-hoc correlation analysis showed that this was not explained by the likeability ratings of the neutral voice (see supplementary materials S4). Rather, this supports the notion that MACH interpersonal styles might make use of nonverbal vocal strategies to make favourable impressions on others. Interestingly, this suggests that strategic social behaviour is not generally higher in socially opportunistic personality styles but rather particularly related to achieving favourable impressions in correspondence within a given situational context. This adds to the findings that Machiavellianism is associated with increased vocal convergence to interlocutors, given that it is advantageous for the speaker (Muir et al., 2016). Muir et al. (2016) showed that both high levels of Machiavellianism and self-monitoring were associated with an increased likelihood of converging linguistically to an interlocutor, though only if the speaker was in a position of lower social power and dependent on making a favourable impression. We also extend previous work by showing that such strategic nonverbal behaviour might include volitional voice changes alongside volitional facial expression (Book et al., 2015). This type of charming behaviour might facilitate trust and liking - and ultimately, rapport - in an interaction and thus contribute to the success with which Machiavellists manipulate others (Jonason & Webster, 2012; Paulhus & Williams, 2002). Better performance in volitional voice modulation might be achieved through increased self-monitoring (Corral & Calvete, 2000; Grieve, 2011) and higher emotional self-control (Nagler et al., 2014), which might support volitional display of emotion in nonverbal behaviours ("posing"; reviewed by DePaulo, 1992).

Differential functional activation associated with social reactivity. We computed a multiple regression analysis to unveil possible functional processing mechanisms contributing to better performance in likeable voice modulation in association with higher levels of Machiavellianism. We found that MACH traits were not uniquely associated with functional processing during likeable voice production on a whole-brain level. This might imply that other processing stages might be associated with MACH traits, for instance during preparation for voice production, or that MACH traits did not uniquely contribute to the model in relation to empathic indices. We therefore conducted post-hoc exploratory analyses of the multiple regression model but this time only including the MACH subscale and gender and age as covariates of no interest (see supplementary materials S6). Here we found a significant negative association between trait MACH and task-related functional processing in two small clusters with peak voxels in bilateral precuneus and a significant positive association with activation in left premotor cortex. The cluster in premotor cortex overlapped with a sole cluster associated with likeability performance (Δ-Likeability), suggesting MACH traits might engage increasingly in precise motor control than social processing during social voice modulations. However, future work is needed to investigate whether the vocal performance to express likeability is associated with changed functional processing during other processing stages.

While MACH traits were positively related to the outcome of the vocal control task, the neural model only revealed unique associations with individual differences in empathy and psychopathy. This implies that the task of sounding likeable might be tackled differently on the neural level regardless of its behavioural success in relation to empathy components. Spatially separate social processing regions were associated with functional activation during likeable voice modulation in opposite directions. Previous work has suggested that overlapping but separable neural regions might support affective and cognitive social processing (Kanske, Böckler, Trautwein, Lesemann, & Singer, 2016; Shamay-Tsoory, Aharon-Peretz, & Perry, 2009; Tholen, Trautwein, Böckler, Singer, & Kanske, 2020; Völlm et al., 2006). In our experiment, higher levels of affective empathy were positively associated with increased activation in precuneus, ventral mPFC and left pTPJ during the production of likeable speech. On the other hand, lower levels of cognitive empathy were associated with higher activation in dmPFC, right STS and left parahippocampus and right orbital IFG. Right orbital IFG activation was also negatively associated with psychopathic traits. The observed regions overlap largely with a set of regions showing increased functional connectivity during life-conversation in relation to social impairment in autistic speakers compared to neurotypicals (Jasmin et al., 2019) specifically between precuneus, right paraHC, right IFG. As both groups of speakers were matched in task performance, the authors suggest compensatory processing in these regions to support social interaction. Our findings add that this might also be the case for nonverbal social vocal behaviour in relation to empathy traits in normal speakers. However, given that the differences in neural processing observed in relation to empathy indices and psychopathy were not reflected in better voice modulation performance, and together with the supplementary findings in relation to MACH (supplementary materials S6), this might imply that particularly motor-planning and implementation supported by premotor cortices is necessary to successful volitional voice modulation, at least in isolation, while different (possibly compensatory) strategies might support the expression of social information in the voice.

Considerations and further studies

Onew limitation of this study is that our sample included a majority of female speakers and raters. Previous work has shown gender differences in some acoustic parameters in response to expressing social traits in the voice (Hughes et al., 2014), although these might be more important in mating contexts, which often rely on exaggerations of sexually dimorphic physical characteristics (Fraccaro et al., 2013; Jones, Feinberg, DeBruine, Little, & Vukovic, 2010). Although this is not to be expected for the traits we studied here (intelligence, likability and hostility), we controlled statistically for effects of gender in all of the reported analyses. Future work is needed with a more balanced sample, to investigate possible gender differences in social vocal control ability. Another limitation is the reliance on self-report measures: although our instruments showed satisfactory internal consistencies and variation comparable to previous work in non-clinical populations (SD3: Jones & Paulhus, 2014; Maples, Lamkin, & Miller, 2014; QCAE: Seara-Cardoso, Dolberg, Neumann, Roiser, & Viding, 2013), the SD3 has recently come under scrutiny (Persson, Kajonius, & Garcia, 2017). Behavioural measures of cognitive and affective empathy components (e.g. EmpaToM; Kanske, Böckler, Trautwein, & Singer, 2015) would be useful in future work.

The theoretical RSM was based on perceptual organisation of both neutral and modulated voice samples in the social voice space (McAleer et al., 2014), which assumes that likeability and hostility are at the two opposite poles of an affiliation (warmth/trustworthiness) dimension, whereas intelligence is related to an orthogonal competence (dominance/hierarchy) dimension. Thus, we assumed that maximal specificity of trait expression would lead to naïve listeners' ratings of likeability and hostility to be anti-correlated, whereas there should be no significant association between intelligence ratings with the other traits. One could imagine alternative RSM patterns, e.g. with a higher correlation between more socially favourable traits (intelligence and likeability) in contrast to aversive traits (hostility). However, the perceptual organisation of voice in the two-dimensional social voice space we modelled here is robust across different languages and stimulus materials (Baus, McAleer, Marcoux, Belin, & Costa, 2019; Mahrholz et al., 2018), and therefore offers a good starting point. Future work could explore different RSM patterns, particularly when investigating underlying neural processing strategies.

Here, we report evidence that social reactivity-related personality traits might account for differences in functional processing to achieve voluntary social voice changes, although we did not find evidence of an association between social vocal control ability and social reactivity on the behavioural level. Previous work shows that clinical populations with pervasive social impairments, such as ASD and psychopathy, also show deficits in the perception (Bagley, Abramowitz, & Kosson, 2009; Blair et al., 2002; Globerson, Amir, Kishon-Rabin, & Golan, 2015) and spontaneous production of socio-emotional information in the voice (Fusaroli et al., 2016; Louth et al., 1998; McCann & Peppé, 2003; Shriberg et al., 2001; ten Brinke et al., 2017). Our findings in a subclinical sample suggest that although spontaneous vocal behaviour might be similarly impaired in both disorders, voluntary social behaviour might be specifically impaired in ASD, but not in psychopathy. More work with clinical populations, preferably using an endophenotypic transdiagnostic approach (Insel et al., 2010), is needed to disentangle social vocal control in ASD and Psychopathy. Moreover, future studies should therefore take the distinction between spontaneous vocal expression and voluntary vocal expression into account. In fact, existing training programs to increase nonverbal vocal communicatory abilities in ASD (e.g. Dunn, Harris, & Dunn, 2017), already make use of volitional voice changes to support treatment.

Conclusion

Our findings suggest that individual differences in social vocal control are not generally related to social reactivity, but that performance in expressing favourable traits (sounding likeable) is specifically associated with dispositional Machiavellianism. Based on our observations, we conclude that social opportunists might engage in goal-directed vocal behaviour to navigate social interactions. Further, we find that the efficacy with which social traits are volitionally expressed in the voice relies on a network of regions supporting fine-tuned motor planning, auditory and somatosensory feedback control and socio-cognitive processing, and that differential engagement of these regions is associated with a speaker's dispositional cognitive and affective empathy.

CRediT author statement. Stella Guldner: Conceptualization, methodology, software, investigation, formal analysis, writing – original draft, visualization. Frauke Nees: Conceptualization, writing – review & editing, supervision. Herta Flor: Supervision, writing – review & editing. Carolyn McGettigan: Conceptualization, resources, methodology, supervision, project administration, funding acquisition, writing – review & editing.

Acknowledgement. We thank Elise Kanber for her help with data acquisition.

Conflict of Interest. The authors declare no conflict of interest.

Funding. This work was supported by a Research Leadership Award from The Leverhulme Trust (RL-2016-013) awarded to CM.

References

Amenta, S., Noël, X., Verbanck, P., & Campanella, S. (2013). Decoding of Emotional Components in Complex Communicative Situations (Irony) and Its Relation to Empathic Abilities in Male Chronic Alcoholics: An Issue for Treatment. *Alcoholism: Clinical and Experimental Research*, *37*(2), 339–347. https://doi.org/10.1111/j.1530-0277.2012.01909.x

Austin, E. J., Farrelly, D., Black, C., & Moore, H. (2007). Emotional intelligence, Machiavellianism and emotional manipulation: Does EI have a dark side? *Personality and Individual Differences*, *43*(1), 179–189. https://doi.org/10.1016/j.paid.2006.11.019

Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS ONE*, *5*(1), e8759. https://doi.org/10.1371/journal.pone.0008759

Bagley, A. D., Abramowitz, C. S., & Kosson, D. S. (2009). Vocal affect recognition and psychopathy: converging findings across traditional and cluster analytic approaches to assessing the construct. *Journal of Abnormal Psychology*, *118*(2), 388–398. https://doi.org/10.1037/a0015372

Baus, C., McAleer, P., Marcoux, K., Belin, P., & Costa, A. (2019). Forming social impressions from voices in native and foreign languages. *Scientific Reports*, *9*(1), 414. https://doi.org/10.1038/s41598-018-36518-6

Benjamini, Y., & Hochberg, Y. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300. https://doi.org/10.1111/j.2517-6161.1995.tb02031.x

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews*, *47*, 520–532. https://doi.org/10.1016/j.neubiorev.2014.09.021

Blair, R. J.R. (2005). Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and Cognition*, *14*(4), 698–718. https://doi.org/10.1016/j.concog.2005.06.004

Blair, R James R, Mitchell, D. G. V, Richell, R. A., Kelly, S., Leonard, A., Newman, C., & Scott, S. K. (2002). Turning a deaf ear to fear: impaired recognition of vocal affect in psychopathic individuals. *Journal of Abnormal Psychology*, *111*(4), 682–686.

Book, A., Methot, T., Gauthier, N., Hosker-Field, A., Forth, A., Quinsey, V., & Molnar, D. (2015). The mask of sanity revisited: Psychopathic traits and affective mimicry. *Evolutionary Psychological Science*, *1*(2), 91–102. https://doi.org/10.1007/s40806-015-0012-x

Brown, S., Cockett, P., & Yuan, Y. (2019). The neuroscience of Romeo and Juliet: An fMRI study of acting. *Royal Society Open Science*, *6*(3), 181908. https://doi.org/10.1098/rsos.181908

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What's New, Pussycat? OnTalking to Babies and Animals. *Science*, *296*(5572), 1435. https://doi.org/10.1126/science.1069587

Bzdok, D., Langner, R., Schilbach, L., Jakobs, O., Roski, C., Caspers, S., … Eickhoff, S. B. (2013). Characterization of the temporo-parietal junction by combining data-driven parcellation, complementary connectivity analyses, and functional decoding. *NeuroImage*, *81*, 381–392. https://doi.org/10.1016/j.neuroimage.2013.05.046

Bzdok, D., Schilbach, L., Vogeley, K., Schneider, K., Laird, A. R., Langner, R., & Eickhoff, S. B. (2012). Parsing the neural correlates of moral cognition: ALE meta-analysis on morality, theory of mind, and empathy. *Brain Structure and Function*, *217*(4), 783–796. https://doi.org/10.1007/s00429-012-0380-y

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends in Cognitive Sciences*, *16*(6), 338–352. https://doi.org/10.1016/j.tics.2012.04.008

Carter, R. M. K., & Huettel, S. A. (2013). A nexus model of the temporal-parietal junction. *Trends in Cognitive Sciences*, *17*(7), 328–336. https://doi.org/10.1016/j.tics.2013.05.007

Ciaramelli, E., Grady, C., Levine, B., Ween, J., & Moscovitch, M. (2010). Top-down and bottom-up attention to memory are dissociated in posterior parietal cortex: Neuroimaging and neuropsychological evidence. *Journal of Neuroscience*, *30*(14), 4943–4956. https://doi.org/10.1523/JNEUROSCI.1209-09.2010

Corral, S., & Calvete, E. (2000). Machiavellianism: Dimensionality of the Mach IV and its Relation to Self-Monitoring in a Spanish Sample. *Spanish Journal of Psychology*, *3*(1), 3–13. https://doi.org/10.1017/S1138741600005497

Cuddy, A. J. C., Fiske, S. T., & Glick, P. (2008). Warmth and Competence as Universal Dimensions of Social Perception: The Stereotype Content Model and the BIAS Map. *Advances in Experimental Social Psychology*, *40*(07), 61–149. https://doi.org/10.1016/S0065-2601(07)00002-0

DePaulo, B. M. (1992). Nonverbal Behavior and Self-Presentation. *Psychological Bulletin*, *111*(2), 203–243.

Dunn, M. A., Harris, L., & Dunn, J. (2017). *Prosody Intervention for High-Functioning Adolescents and Adults with Autism Spectrum Disorder : Enhancing Communication and Social Engagement Through Voice, Rhythm, and Pitch*. London and Philadelphia: Jessica Kingsley Publishers.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton & Co.

Farley, S. D., Hughes, S. M., & LaFayette, J. N. (2013). People Will Know We Are in Love: Evidence of Differences Between Vocal Samples Directed Toward Lovers and Friends. *Journal of Nonverbal Behavior*, *37*(3), 123–138. https://doi.org/10.1007/s10919-013-0151-3

Fehr, B., Samson, D., & Paulhus, D. L. (1992). The construct of Machiavellianism: Twenty years later. In *Advances in personality assessment, Vol. 9.* (pp. 77–116). Hillsdale, NJ, US: Lawrence Erlbaum Associates, Inc.

Fiske, S. T., Cuddy, A. J. C., & Glick, P. (2007). Universal dimensions of social cognition: warmth and competence. *Trends in Cognitive Sciences*, *11*(2), 77–83. https://doi.org/10.1016/j.tics.2006.11.005

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902. https://doi.org/10.1037/0022-3514.82.6.878

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. https://doi.org/10.1016/j.anbehav.2012.10.016

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218

Frühholz, S., Klaas, H. S., Patel, S., & Grandjean, D. (2015). Talking in fury: The cortico-subcortical network underlying angry vocalizations. *Cerebral Cortex*, *25*(9), 2752–2762. https://doi.org/10.1093/cercor/bhu074

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, *10*(3), 384–407. https://doi.org/10.1002/aur.1678

Globerson, E., Amir, N., Kishon-Rabin, L., & Golan, O. (2015). Prosody Recognition in

Adults With High-Functioning Autism Spectrum Disorders: From Psychoacoustics to Cognition. *Autism Research*, *8*(2), 153–163. https://doi.org/10.1002/aur.1432

Golfinopoulos, E., Tourville, J. A., Bohland, J. W., Ghosh, S. S., Nieto-Castanon, A., & Guenther, F. H. (2011). FMRI investigation of unexpected somatosensory feedback perturbation during speech. *NeuroImage*, *55*(3), 1324–1338. https://doi.org/10.1016/j.neuroimage.2010.12.065

Grieve, R. (2011). Mirror Mirror: The role of self-monitoring and sincerity in emotional manipulation. *Personality and Individual Differences*, *51*(8), 981–985. https://doi.org/10.1016/j.paid.2011.08.004

Guenther, F. H. (2006). Cortical interactions underlying the production of speech sounds. *Journal of Communication Disorders*, *39*(5), 350–365. https://doi.org/10.1016/j.jcomdis.2006.06.013

Guldner, S., Nees, F., & McGettigan, C. (2020). Vocomotor and Social Brain Networks Work Together to Express Social Traits in Voices. *Cerebral Cortex*, *30*(11), 6004–6020. https://doi.org/10.1093/cercor/bhaa175

Hallam, G. P., Webb, T. L., Sheeran, P., Miles, E., Niven, K., Wilkinson, I. D., … Farrow, T. F. D. (2014). The neural correlates of regulating another person's emotions: An exploratory fMRI study. *Frontiers in Human Neuroscience*, *8*(376). https://doi.org/10.3389/fnhum.2014.00376

Hellbernd, N., & Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Social Cognitive and Affective Neuroscience*, *13*(6), 604–615. https://doi.org/10.1093/scan/nsy034

Hughes, S. M., & Miller, N. E. (2016). What sounds beautiful looks beautiful stereotype: The matching of attractiveness of voices and faces. *Journal of Social and Personal Relationships*, *33*(7), 984–996. https://doi.org/10.1177/0265407515612445

Hughes, S. M., Mogilski, J. K., & Harrison, M. A. (2014). The Perception and Parameters of Intentional Voice Manipulation. *Journal of Nonverbal Behavior*, *38*(1), 107–127. https://doi.org/10.1007/s10919-013-0163-z

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Jacob, H., Kreifelts, B., Nizielski, S., Schütz, A., & Wildgruber, D. (2016). Effects of emotional intelligence on the impression of irony created by the mismatch between verbal and nonverbal cues. *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0163211

Jasmin, K., Gotts, S. J., Xu, Y., Liu, S., Riddell, C. D., Ingeholm, J. E., … Martin, A. (2019). Overt social interaction and resting state in young adult males with autism: Core and contextual neural features. *Brain*, *142*(3), 808–822. https://doi.org/10.1093/brain/awz003

Jonason, P. K., & Krause, L. (2013). The emotional deficits associated with the Dark Triad traits: Cognitive empathy, affective empathy, and alexithymia. *Personality and Individual Differences*, *55*(5), 532–537. https://doi.org/10.1016/j.paid.2013.04.027

Jonason, P. K., & Webster, G. D. (2012). A protean approach to social influence: Dark Triad personalities and social influence tactics. *Personality and Individual Differences*, *52*(4), 521–526. https://doi.org/10.1016/j.paid.2011.11.023

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. https://doi.org/10.1016/j.anbehav.2009.10.003

Jones, D. N., & Paulhus, D. L. (2014). Introducing the Short Dark Triad (SD3): A Brief Measure of Dark Personality Traits. *Assessment*, *21*(1), 28–41. https://doi.org/10.1177/1073191113514105

Jones, D., & Paulhus, D. (2009). Machiavellianism. Guilford Press.

Kanske, P., Böckler, A., Trautwein, F. M., Lesemann, F. H. P., & Singer, T. (2016). Are strong empathizers better mentalizers? Evidence for independence and interaction between the routes of social cognition. *Social Cognitive and Affective Neuroscience*, *11*(9), 1383–1392. https://doi.org/10.1093/scan/nsw052

Kanske, P., Böckler, A., Trautwein, F. M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and Theory of Mind. *NeuroImage*, *122*, 6–19. https://doi.org/10.1016/j.neuroimage.2015.07.082

Kemper, S., Finter-Urczyk, A., Ferrell, P., Harden, T., & Billington, C. (1998). Using elderspeak with older adults. *Discourse Processes*, *25*(1), 55–73. https://doi.org/10.1080/01638539809545020

Klasen, M., von Marschall, C., Isman, G., Zvyagintsev, M., Gur, R. C., & Mathiak, K. (2018). Prosody production networks are modulated by sensory cues and social context. *Social Cognitive and Affective Neuroscience*, *13*(4), 418–429. https://doi.org/10.1093/scan/nsy015

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Krueger, F., Barbey, A. K., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*(3), 103–109. https://doi.org/10.1016/j.tics.2008.12.005

Kuhn, L. K., Wydell, T., Lavan, N., McGettigan, C., & Garrido, L. (2017). Similar representations of emotions across faces and voices. *Emotion*, *17*(6), 912–937. https://doi.org/10.1037/emo0000282

Lockwood, P. L., Bird, G., Bridge, M., & Viding, E. (2013). Dissecting empathy: high levels of psychopathic and autistic traits are characterized by difficulties in different social information processing domains. *Frontiers in Human Neuroscience*, *7*(760). https://doi.org/10.3389/fnhum.2013.00760

Lombardo, M. V., Chakrabarti, B., Bullmore, E. T., Wheelwright, S. J., Sadek, S. A., Suckling, J., … Williams, S. C. (2010). Shared neural circuits for mentalizing about the self and others. *Journal of Cognitive Neuroscience*, *22*(7), 1623–1635. https://doi.org/10.1162/jocn.2009.21287

Louth, S. M., Williamson, S., Alpert, M., Pouget, E. R., & Hare, R. D. (1998). Acoustic distinctions in the speech of male psychopaths. *Journal of Psycholinguistic Research*, *27*(3), 375–384. https://doi.org/10.1023/A:1023207821867

Ma, N., Baetens, K., Vandekerckhove, M., Kestemont, J., Fias, W., & Van Overwalle, F. (2014). Traits are represented in the medial prefrontal cortex: An fMRI adaptation study. *Social Cognitive and Affective Neuroscience*, *9*(8), 1185–1192. https://doi.org/10.1093/scan/nst098

Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker’s personality are correlated across differing content and stimulus type. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0204991

Maples, J. L., Lamkin, J., & Miller, J. D. (2014). A test of two brief measures of the dark triad: The dirty dozen and short dark triad. *Psychological Assessment*, *26*(1), 326–331. https://doi.org/10.1037/a0035084

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say “hello”? Personality impressions from brief novel voices. *PLoS ONE*, *9*(3), e90779. https://doi.org/10.1371/journal.pone.0090779

McCann, J., & Peppé, S. (2003). Prosody in autism spectrum disorders: A critical review. *International Journal of Language and Communication Disorders*, *38*(4), 325–350. https://doi.org/10.1080/1368282031000154204

McGettigan, C. (2015). The social life of voices: studying the neural bases for the expression and perception of the self and others during spoken communication. *Frontiers in Human Neuroscience*, *9*(129). https://doi.org/10.3389/fnhum.2015.00129

McGettigan, C., Eisner, F., Agnew, Z. K., Manly, T., Wisbey, D., & Scott, S. K. (2013). T'ain't what you say, it's the way that you say it--left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. *Journal of Cognitive Neuroscience*, *25*(11), 1875–1886. https://doi.org/10.1162/jocn_a_00427

Molenberghs, P., Johnson, H., Henry, J. D., & Mattingley, J. B. (2016). Understanding the minds of others: A neuroimaging meta-analysis. *Neuroscience and Biobehavioral Reviews*, *65*, 276–291. https://doi.org/10.1016/j.neubiorev.2016.03.020

Mondino, M., Poulet, E., Suaud-Chagny, M. F., & Brunelin, J. (2016). Anodal tDCS targeting the left temporo-parietal junction disrupts verbal reality-monitoring. *Neuropsychologia*, *89*, 478–484. https://doi.org/10.1016/j.neuropsychologia.2016.07.022

Montepare, J. M., & Zebrowitz-McArthur, L. (1987). Perceptions of adults with childlike voices in two cultures. *Journal of Experimental Social Psychology*, *23*(4), 331–349. https://doi.org/10.1016/0022-1031(87)90045-X

Muir, K., Joinson, A., Cotterill, R., & Dewdney, N. (2016). Characterizing the Linguistic Chameleon: Personal and Social Correlates of Linguistic Style Accommodation. *Human Communication Research*, *42*(3), 462–484. https://doi.org/10.1111/hcre.12083

Nagler, U. K. J., Reiter, K. J., Furtner, M. R., & Rauthmann, J. F. (2014). Is there a "dark intelligence"? Emotional intelligence is used by dark personalities to emotionally manipulate others. *Personality and Individual Differences*, *65*, 47–52. https://doi.org/10.1016/j.paid.2014.01.025

Neves, L., Cordeiro, C., Scott, S. K., Castro, S. L., & Lima, C. F. (2018). High emotional contagion and empathy are associated with enhanced detection of emotional authenticity in laughter. *Quarterly Journal of Experimental Psychology*, *71*(11), 2355–2363. https://doi.org/10.1177/1747021817741800

Nicolle, A., Klein-Flügge, M. C., Hunt, L. T., Vlaev, I., Dolan, R. J., & Behrens, T. E. J. (2012). An Agent Independent Axis for Executed and Modeled Choice in Medial

Prefrontal Cortex. *Neuron*, *75*(6), 1114–1121. https://doi.org/10.1016/j.neuron.2012.07.023

O'Connor, J. J. M., & Barclay, P. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, *38*(4), 506–512. https://doi.org/10.1016/j.evolhumbehav.2017.03.001

Oberhuber, M., Hope, T. M. H., Seghier, M. L., Parker Jones, O., Prejawa, S., Green, D. W., & Price, C. J. (2016). Four Functionally Distinct Regions in the Left Supramarginal Gyrus Support Word Processing. *Cerebral Cortex*, *26*(11), 4212–4226. https://doi.org/10.1093/cercor/bhw251

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Pavela Banai, I., Banai, B., & Bovan, K. (2017). Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evolution and Human Behavior*, *38*(3), 309–314. https://doi.org/10.1016/j.evolhumbehav.2016.10.012

Persson, B. N., Kajonius, P. J., & Garcia, D. (2019). Revisiting the Structure of the Short Dark Triad. *Assessment*, *26*(1), 3–16. https://doi.org/10.1177/1073191117701192

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice Modulation: A Window into the Origins of Human Vocal Control? *Trends in Cognitive Sciences*, *20*(4), 304–318. https://doi.org/10.1016/j.tics.2016.01.002

Pisanski, K., Mora, E. C., Pisanski, A., Reby, D., Sorokowski, P., Frackowiak, T., & Feinberg, D. R. (2016). Volitional exaggeration of body size through fundamental and formant frequency modulation in humans. *Scientific Reports*, *6*, 34389. https://doi.org/10.1038/srep34389

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, *63*(11), 2251–2272. https://doi.org/10.1080/17470211003721642

Saxe, R., & Wexler, A. (2005). Making sense of another mind: The role of the right temporo-parietal junction. *Neuropsychologia*, *43*(10), 1391–1399. https://doi.org/10.1016/j.neuropsychologia.2005.02.013

Schroeder, J., & Epley, N. (2015). The Sound of Intellect: Speech Reveals a Thoughtful Mind, Increasing a Job Candidate's Appeal. *Psychological Science*, *26*(6), 877–891. https://doi.org/10.1177/0956797615572906

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Seara-Cardoso, A., Dolberg, H., Neumann, C., Roiser, J. P., & Viding, E. (2013). Empathy, morality and psychopathic traits in women. *Personality and Individual Differences*, *55*(3), 328–333. https://doi.org/10.1016/j.paid.2013.03.011

Shamay-Tsoory, S. G., Aharon-Peretz, J., & Perry, D. (2009). Two systems for empathy: a double dissociation between emotional and cognitive empathy in inferior frontal gyrus versus ventromedial prefrontal lesions. *Brain*, *132*(3), 617–627. https://doi.org/10.1093/brain/awn279

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A. M., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research : JSLHR*, *44*(5), 1097–1115.

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *Neuroscientist*, *17*(2), 197–208. https://doi.org/10.1177/1073858410386727

Slotnick, S. D., Moo, L. R., Segal, J. B., & Hart, J. (2003). Distinct prefrontal cortex activity associated with item memory and source memory for visual shapes. *Cognitive Brain Research*, *17*(1), 75–82. https://doi.org/10.1016/S0926-6410(03)00082-X

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 117–137). Oxford Univ. Press.

Szabó, E., & Bereczkei, T. (2017). Different paths to different strategies? Unique associations among facets of the Dark Triad, empathy, and trait emotional intelligence. *Advances in Cognitive Psychology*, *13*(4), 306–313. https://doi.org/10.5709/acp-0230-7

ten Brinke, L., Porter, S., Korva, N., Fowler, K., Lilienfeld, S. O., & Patrick, C. J. (2017). An Examination of the Communication Styles Associated with Psychopathy and Their Influence on Observer Impressions. *Journal of Nonverbal Behavior*, *41*(3), 269–287. https://doi.org/10.1007/s10919-017-0252-5

Tholen, M. G., Trautwein, F. M., Böckler, A., Singer, T., & Kanske, P. (2020). Functional magnetic resonance imaging (fMRI) item analysis of empathy and theory of mind. *Human Brain Mapping*, *41*(10), 2611–2628. https://doi.org/10.1002/hbm.24966

Thornton, X. M. A., Weaverdyck, M. E., & Tamir, X. I. (2019). The social brain automatically predicts others' future mental states. *Journal of Neuroscience*, *39*(1), 140–148. https://doi.org/10.1523/JNEUROSCI.1431-18.2018

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, *33*(3), 210–216. https://doi.org/10.1016/j.evolhumbehav.2011.09.004

Torre, I., Goslin, J., White, L., & Zanatto, D. (2018). Trust in artificial voices. In *Proceedings of the Technology, Mind, and Society on ZZZ - TechMindSociety '18* (Vol. 18, pp. 1–6). https://doi.org/10.1145/3183654.3183691

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Van Overwalle, F., Ma, N., & Baetens, K. (2016). Nice or nerdy? The neural representation of social and competence traits. *Social Neuroscience*, *11*(6), 567–578. https://doi.org/10.1080/17470919.2015.1120239

Völlm, B. A., Taylor, A. N. W., Richardson, P., Corcoran, R., Stirling, J., McKie, S., … Elliott, R. (2006). Neuronal correlates of theory of mind and empathy: A functional magnetic resonance imaging study in a nonverbal task. *NeuroImage*, *29*(1), 90–98. https://doi.org/10.1016/j.neuroimage.2005.07.022

Wai, M., & Tiliopoulos, N. (2012). The affective and cognitive empathic nature of the dark triad of personality. *Personality and Individual Differences*, *52*(7), 794–799. https://doi.org/10.1016/j.paid.2012.01.008

Wang, H., Callaghan, E., Gooding-Williams, G., Mcallister, C., & Kessler, K. (2016). Rhythm makes the world go round: An MEG-TMS study on the role of right TPJ theta oscillations in embodied perspective taking. *CORTEX*, *75*, 68–81. https://doi.org/10.1016/j.cortex.2015.11.011

Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, *13*(2), 67–82. https://doi.org/10.1007/BF00990791

Supplementary Materials

S1. Acquisition of Imaging Data. All functional brain images were recorded on a 3T Siemens TIM Trio scanner with a 32 channel headcoil, using a rapid–sparse event-related 3D echo-planar imaging (EPI) sequence (32 axial slices, slice gap 25%, resolution 3x3x3mm2, flip angle 78°, matrix 64 x 64, TE: 30 msec, TR: 3.5 sec, TA: 2 sec). A 3D T1-weighted MP-RAGE scan was acquired for EPI image alignment and spatial normalisation (voxel size 1 mm isotropic; flip angle 11°; TE 3.03 ms; TR 1830 ms; image matrix 256 x 256). Analysis was conducted in SPM12 (http://www.fil.ion.ucl.ac.uk/spm/). Preprocessing steps included spatial realignment, segmentation, co-registration, normalization (functional images were resampled to a voxel size of 2x2x2mm) and smoothing (FWHM=8mm). 1st Level general linear models included the vocal control task conditions as regressors and subjective ratings as parametric modulators for each condition.

S2. Psychophysiological Interactions. Each speaker's individual core vocal motor area was identified on the basis of their peak activation during the Go vs No-Go contrast within a VOI (voxels of interest; 10 mm radius) around the group-level peak activation of the same contrast in left IFG [-52 12 24] at a voxel height threshold of p<.001. 6 mm spheres were constructed around this individual peak activations and from this we extracted the first eigenvariate of the functional MRI signal change. The mean time course was then multiplied by the task regressors based on the Go > No-Go Modulation contrast. We added this interaction term as regressors to the 1st level models, along with the deconvolved source signal of the VOI.

S3. Exploratory functional activations for multiple regression model of Δ-Likeability during Likeable Go trials. Only positive correlations survived.

Table S3. Functional activations associated with performance in likeable voice modulations (Δ-Likeability) during Likeable Go > Rest.

Contrast	***k***	**Region**	**Hem.**	**Coordinate**			***T***	***Z***
				x	*y*	*z*		
Likeable performance on Likeable go > rest	61	Middle frontal gyrus	L	-34	10	42	5.48	4.23

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. *p*<.001 uncorrected, minimal cluster size: 60 voxels.

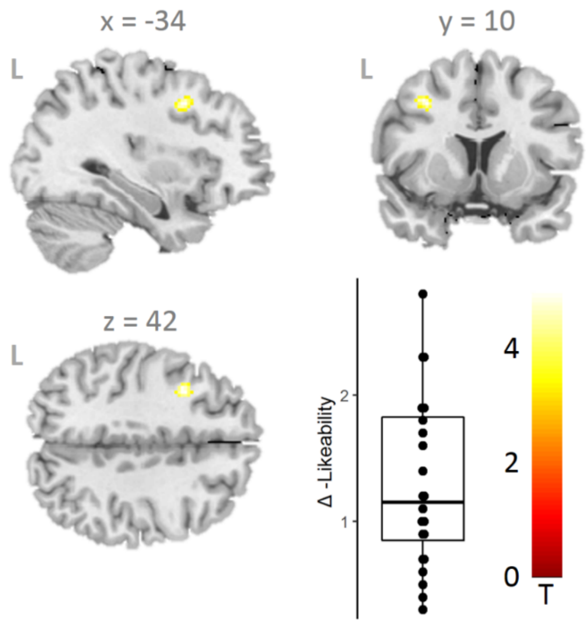
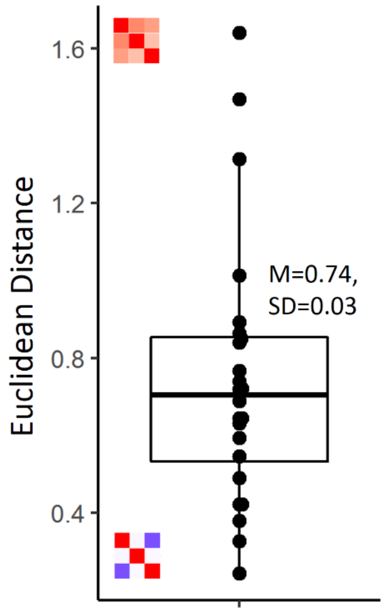


Figure S3. Activation Maps for Likeability Performance on Likeable Go > Rest and descriptive stats of Likeable voice performance. Performance in likeable voice modulation was positively associated with functional activation in a cluster in middle frontal gyrus. L=left.

Figure S4. General social vocal control performance. Higher values indicate worse performance in social vocal control ability, operationalized to worse specificity in evoked trait percepts in listeners. Exemplary RSMs for two speakers are given to illustrate better and worse specificity of evoked trait percepts, reflected in pairwise correlation coefficients in each cell. The minimum of the Euclidean distance measure would be 0, assuming that a speaker achieves maximal differentiation between evoked trait ratings (the speaker's RSM and the theoretical RSM would be identical), whereas 2.45 would be the maximum distance (if the speaker's RSM would be a matrix of ones, i.e. no differentiation between the trait categories). Descriptive statistics are reported in the main text.



S5. Correlations between naïve ratings of likeability in neutral voices and MACH. To test for possible baseline effects of MACH on perception of likeability of speakers' neutral voices, we did additional post-hoc correlation analysis to compare mean likeability ratings of neutral voices and speakers' MACH scores. Partial Pearson correlation analysis revealed no significant association between MACH and likeability ratings of neutral voices (r_p=-.03, p=.89) controlling for sex and age.

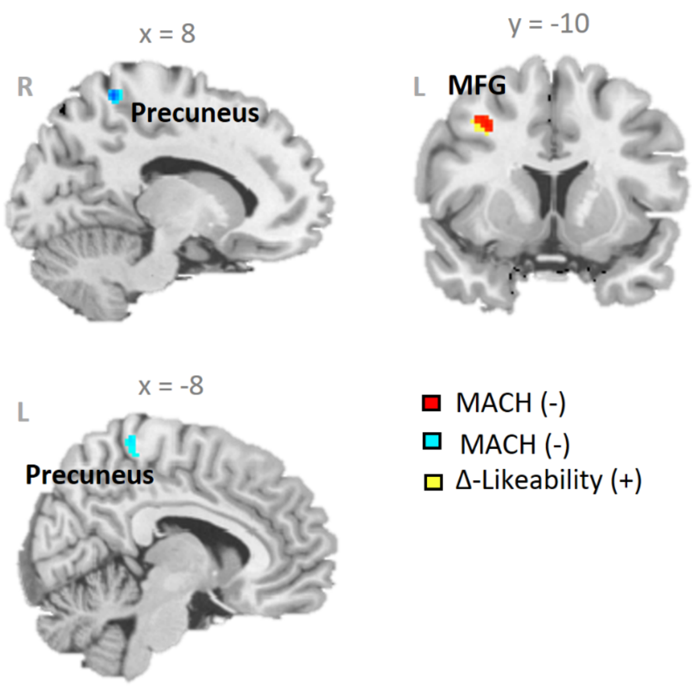
S6. Exploratory functional activations for multiple regression model of MACH during Likeable Go trials. The multiple regression model showed a negative association with MACH in bilateral precuneus (blue) and a positive association with activation in middle frontal gyrus during likeable voice modualtions (red; uncorrected p<.001, k=25). The cluster in MFG overlapped with a cluster associated with likeable voice performance (Δ-Likeability; see S3).

Table S6. Functional activations associated with Machiavellianism during Likeable Go > Rest.

Contrast	*k*	Region	Hem.	Coordinate *x*	*y*	*z*	*T*	*Z*
MACH (-)	50	Precuneus	L	-10	-40	66	5.90	4.44
	44	Precuneus	R	12	-48	64	4.33	3.59
MACH (+)	43	Middle frontal gyrus	L	-34	10	42	4.66	3.79

Note. *k*, cluster size in number of voxels, Hem., Hemisphere, L, left, R, right. Coordinates are in Montreal Neurological Institute (MNI) stereotactic space. *p*<.001 uncorrected, minimal cluster size: 25 voxels.

Figure S6. Activation Maps for self-report MACH and Likeability Performance on Likeable Go > Rest and descriptive stats of Likeable voice performance. MACH was positively associated with functional activation in a cluster in middle frontal gyrus (red), overlapping with the left MFG cluster related to likeability performance (yellow). MACH was negatively associated with activation in bilateral precuneus (blue). (+/-) denotes positive or negative relationships, respectively. L=left.



3 General Discussion

The goal of this dissertation was to investigate psychological and neurophysiological factors contributing to voluntary social voice modulation. The reported findings offer evidence that social traits can be effectively and specifically expressed in the voice, amplifying perceived social traits within the previously described social voice space. Volitional social voice modulation engages areas associated with domain-general social cognitive processing, which might support the integration of motor plans based on social trait representations during ongoing voice production on the level of a core vocomotor region in left IFG. Activation in left posterior TPJ, premotor cortex and sensorimotor areas was predicted by the specificity with which social voice modulations influenced trait perception of speakers. Thus, the ability with which speakers modulate their voice to express social information seems to be related to both socio-cognitive processing and vocal motor control. Lastly, Machiavellianism, rather than dispositional empathy, was associated with social vocal control ability to evoke favourable social judgements (being perceived as likeable). Below, each of these findings will be reviewed separately. Finally, limitations of this work will be discussed and a conclusion and outlook for further research will be provided.

Functional aspects of Social Vocal Control

This thesis offers evidence that social traits can be effectively and specifically communicated through volitional vocal behaviour both in short sentences and pseudo-words (study 1, 2, 3). This extends previous work on volitional social trait expression in the voice, by using a multivariate rating paradigm to show that speakers can evoke trait perceptions specifically. Previously, Hughes et al. (2014) showed increased trait perception after voice modulation compared to the neutral voice. But the authors only obtained comparisons between the neutral voice and the congruent trait expression. Consequently, it was unclear whether the perception of any voice change leads to an increased trait percept, or whether specific modulations are necessary to express particular traits. Moreover, using RSM analysis (study 3) to measure the specificity of evoked perceptions is a novel approach to voice production research. This type of analysis has been used traditionally to examine mental representation of stimulus categories in listeners (e.g. Kuhn, Wydell, Lavan, McGettigan, & Garrido, 2017). Moreover, we show that volitional voice modulations can amplify the perception of traits within the social trait space (McAleer et al., 2014). This has previously been documented only in artificial modulation of pitch in voices (Belin, Boehme, & McAleer, 2017) and supports the notion that voice changes are a viable tool to manage listeners' impressions of speakers in social interactions. Although listeners' trait judgments do not necessarily reflect reality (Olivola & Todorov, 2010), they are rapid impressions (McAleer et al., 2014) that influence how a speaker is perceived globally – based on a speaker's voice, listeners also infer visual appearance or other personality attributes (Hughes & Miller, 2016; Montepare & Zebrowitz-McArthur, 1987; Zuckerman & Driver, 1989). Although heuristic, these judgements therefore lay ground for a speaker's social attractiveness, underlying the importance to create favourable impressions on others to achieve personal goals in social interactions. Together with trait judgments from faces (Mileva, Tompkinson, Watt, & Burton, 2018), this work therefore advances the current literature by highlighting the importance of vocal behaviour in impression management. Moreover, it underlines the human capacity to engage in goal-directed vocal modulations more broadly as well. Such vocal flexibility is important in social interactions by signalling social distance and affiliation (Giles et al., 1991), increasing efficiency of vocal communication (Neumann & Strack, 2000) and the quality of an interaction (Gregory et al., 1993), both through vocal alignment (imitative processes; Pardo, 2006) and the adaptation to the emotional and social needs of the social context of a speaker (modulatory process; Burnham et al., 2002).

Predictors of social vocal control ability

The work presented here suggests that, in a subclinical population of speakers, neither dispositional empathy (cognitive nor affective aspects; study 3), nor more general vocal control ability (expressing body-size through the voice; study 1) were associated with performance in social vocal control. This was surprising, given that a speaker's dispositional capacity to mentalize about the listener might support adequate social vocal behaviour (Louth et al., 1998; Shriberg et al., 2001) and that affective empathy is related to understanding of emotional (Aziz-Zadeh et al., 2010) and social information from speech (Amenta et al., 2013; Jacob et al., 2016; Neves et al., 2018).

To some extent this might be due to differences in the operationalization of vocal control ability: body-size modulation was measured categorically, while social voice modulation was measured continuously (study 1). Moreover, we measured social vocal control ability as the magnitude of voice change in response to social traits in study 1, but by the more multivariate

Euclidean distance measure in study 3. This was due to the construction of the theoretical RSMs which was based on the inter-correlations between trait ratings. These inter-correlations would have been harder to estimate with multiple traits (e.g. it is unclear how tightly attractiveness and confidence should correlate theoretically), rather than the reduced trait selection (hostility, likeability, intelligence), which allowed a clear theoretical profile based on the social trait space.

Another reason that we have not found significant associations between social reactivity indices and performance in social vocal control might be that social reactivity measures are more important for spontaneous vocal behaviour, particularly in life interaction. Previous work suggests a positive association between empathy measures and social inference from voices (Amenta et al., 2013; Neves et al., 2018), reflected in associated functional processing and connectivity during voice processing (Aziz-Zadeh et al., 2010; Jiang et al., 2017). Thus, empathy might be important to correctly appraise propositional vocal behaviour and emotional needs of others in conversations, to then adapt one's own vocal behaviour accordingly. This is partly supported by the exploratory finding that dispositional empathy might lead to differential neural processing strategies to carry out the task, however, at the same time having no effect on task performance (study 3).

Although it seems unlikely, it could nevertheless be the case that the variability in listeners' rating behaviour might have a deleterious effect on the expected association between social vocal control ability and social reactivity indices. However, the experimental design controlled for differences in raters, as every speaker was rated by a unique set of ten listeners. We refrained from statistically standardizing the obtained ratings (for instance through z-transformations) so as not to compromise the ecological validity of social vocal behaviour by eliminating natural differences in how voices are heard in real-world interactions.

Our findings suggest further that both motor processing and social processing might be involved in social voice modulation on the neural level (study 3). Importantly, socio-emotional expression can be trained, seen for instance, in professional acting. In fact, the effects of training can also be readily observed in vocal artistry during singing or impersonations (Lavan, Burton, et al., 2019). In everyday life, of course, other personality characteristics might also contribute to social vocal control ability, for instance extroversion or self-monitoring. A speaker's level of extroversion is reflected in specific vocal cues

(Scherer, 1978) and might be associated with more opportunities to practice social vocal control. Furthermore, self-monitoring might be associated with higher motivation to express adequate social information in the voice according to the social context (see below, Gangestad & Snyder, 2000). Future work is needed to determine the contribution of training and personality traits on vocal flexibility, beyond those measured and reported here.

Importantly, we show a unique contribution of dispositional Machiavellianism to volitionally expressing favourable traits in the voice (sounding likeable; study 3). This corroborates previous work showing that greater likelihood of vocal alignment is positively associated with Machiavellistic traits if it is beneficial to the speaker (Muir et al., 2016) and suggesting that Machiavellists might strategically target influencing others through tactics such as charming (Jonason & Webster, 2012; Wilson, Near, & Miller, 1996), which is a tactic with low costs and high potential benefits. One route through which this might be achieved is through increased self-monitoring (Corral & Calvete, 2000; Grieve, 2011). Self-monitoring describes a person's ability to control their own socio-emotional expression in order to match it with the current social context depending on own social goals. In fact, people with high levels of self-monitoring are often very good at decoding nonverbal social cues in others in order to adapt their behaviour accordingly (Gangestad & Snyder, 2000). Nonverbal vocal modulation might be part of the self-monitoring repertoire. Importantly, the findings reported here show the unique contribution of Machiavellistic traits over and above psychopathic, cognitive or affective empathy indices, corroborating the very definition of Machiavellistic traits based on social predatory behaviour (Christie & Geis, 1970). Moreover, likeability ratings of neutral voices were not associated with levels of Machiavellianism in speakers, therefore it can be inferred that the success with which likeability is expressed in the voice is related to specific and volitional vocal modulation rather than general expression of socially desirable traits or decreased likeability of the neutral voice. More work is needed to identify these specific changes on the physiological and acoustic level in order to inform neurocognitive models associated with speaking in a charming voice in Machiavellianism.

Neural basis of social vocal control

Regarding the neural mechanisms involved in social vocal control, we report a network particular to volitional voice modulation similar to the previously proposed VMN (insulae,

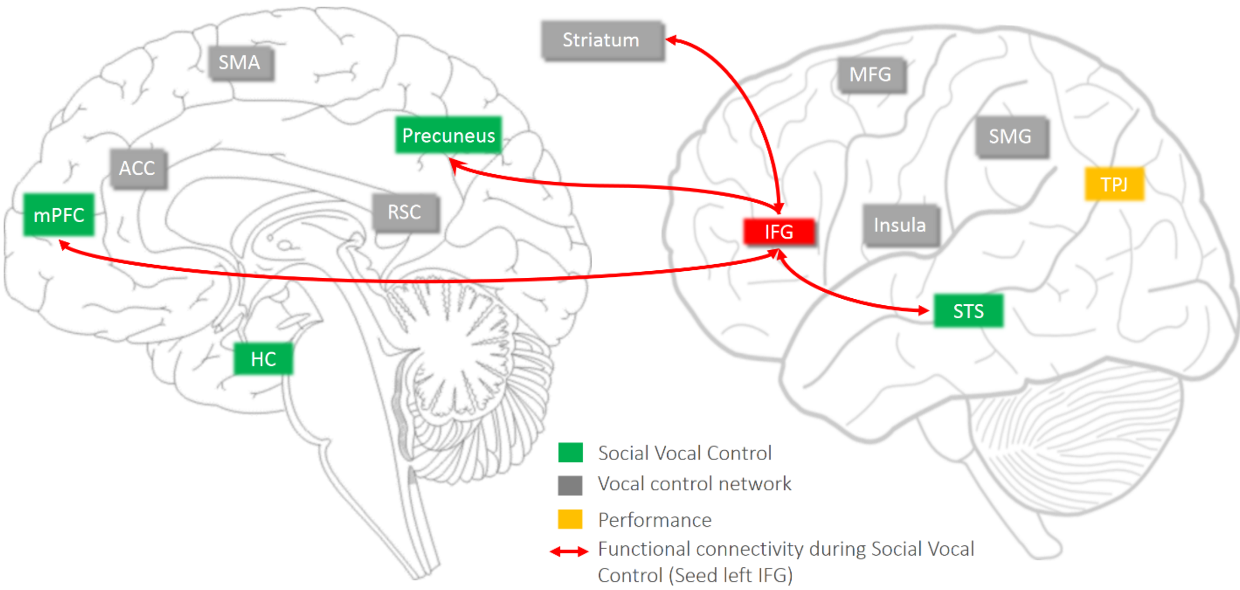


Figure 2. Networks involved in social vocal control. Structures involved in vocomotor control during volitional voice modulation are shown in grey, regions involved in social vocal control in green and areas specifically associated with performance in social voice modulation in yellow. Functional connectivity during social vocal control is depicted in red, with left IFG as source region.

right STC, left IFG, SMA, ACC, SMG and corpus callosum) and add the involvement of precuneus, mPFC and STS in social vocal control. The findings confirm the hypothesis that these regions interact with left IFG to achieve respective social voice changes (study 2). This suggests that activation in regions associated with trait processing and mentalizing might inform speech sound maps represented in left IFG, that are then fed into motor cortex for motor planning and execution (Tourville & Guenther, 2011). Social vocal control might rely on a similar integration of two processing systems as it was proposed for spontaneous (Ackermann et al., 2014) and volitional affective voice modulation (e.g. Frühholz et al., 2015). Individual differences in social vocal control ability revealed that premotor cortex, as well as auditory and somatosensory feedback areas, support performance in vocal control together with activation in left posterior TPJ (pTPJ; study 3). This suggests a possible functional segregation in SBN areas: While mPFC, precuneus and STS support social vocal control possibly via trait processing and speech sound map integration, left pTPJ - involved in social intention inference and explicit mentalizing - supports the performance with which social traits are encoded in the voice. The findings imply that while midline structures and STS might inform motor plans for speech, pTPJ might be specifically involved in monitoring the social content of the speech outcome. Moreover, individual differences in empathy subcomponents might differentially engage social processing streams to social vocal control. Cognitive empathy was negatively associated with task-related functional activation in right STS, orbital parts of IFG, dorsal mPFC, whereas affective empathy components were associated with increased activation in ventral mPFC, precuneus and left pTPJ (study 3). The main findings are summarized in Figure 2.

Open questions remain as to the specific function of mPFC, precuneus, and TPJ related to social emotional processing. From the work reported here, we can speculate that social feedforward and feedback processes might be involved in social vocal control: social trait representation and simulation of the listener's future mental state as a consequence of vocal behaviour might support social vocal control through engagement of mPFC and precuneus (Krueger et al., 2009; Tamir & Thornton, 2018; Thornton, Weaverdyck, & Tamir, 2019), whereas social feedback processing in pTPJ (Hellbernd & Sammler, 2018) might contribute to social vocal control ability. Thus the work presented here offers novel hypotheses for future studies. Beyond advancing our understanding of social processing in the voice, these findings also highlight an important gap in the field of social neurosciences: most studies focus on perceptual processes involved in social processing, but less on active social behaviour. The work reported here suggests structures associated with social perception might also be engaged during socially motivated action, but more work using active social tasks is required. Given the constraints of functional imaging settings regarding movement, studying speech production might offer a good possibility with technical advances allowing for better control of movement related artefacts. Beyond fMRI, functional near-infrared spectroscopy or portable electroencephalography are other techniques allowing for more freedom in movement to study social behaviour and interaction, albeit on more superficial cortical layers.

3.1 Limitations

Findings reported here rely largely on female speakers (study 1: 13/40 male speakers, study 2 and 3: 4/24 male speakers). Given the anatomical differences, and the fact that most previous work has focussed on courtship and mating preferences coming from an evolutionary background, speaker production and listener perception was often contrasted based on gender differences. Although it must be acknowledged that acoustic differences in vocal modulations might be present, this thesis was focused on the effect of social vocal modulations on perception, that is, whether social impressions can be reliably evoked and are therefore effective and impactful for social interactions. In fact, the social inferences drawn from voices might be stable over gender roles (Ponsot, Burred, Belin, & Aucouturier, 2018). Moreover, although acoustic variation is of course a central outcome in voice modulation studies, it also has its limits, often being not able to reliably account for more fine-grained differences in perception (Pardo, 2013). Therefore, naïve listener ratings, with variability in raters across speakers as well as multivariate trait ratings can more precisely capture the social impact of the voice. However, the integration of both naïve rater evoked impressions and acoustic variation should be the way forward in future studies (discussed below).

Although levels of Machiavelliansim were positively associated with a likeable voice percept, it is unclear whether listeners would catch on to the sincerity of the vocal portrayal, which might diminish its favourable social consequences. Authenticity ratings in facial affective mimicry, for instance, were higher in psychopathic participants compared to neurotypicals, although or maybe because the corresponding emotional experience was lacking (Book et al., 2015). In voices, particularly pitch contour is related to emotional authenticity (Anikin & Lima, 2017; Drolet et al., 2014). However, listeners' awareness of inauthentic vocal modulation to express confidence, for instance, has been shown to be independent of the favourable judgments (e.g. of the persuasiveness of the voice; Van Zant & Berger, 2020). Nevertheless, future work should include a measure of perceived authenticity to elucidate whether social vocal control is not only more recognizable with higher levels of Machiavellianism, but also more persuasive. Lastly, behavioural measures of dispositional empathy would be desirable to pinpoint social-cognitive processes involved in volitional vocal behaviour more precisely. One such example would be using EmpaToM (Kanske, Böckler, Trautwein, & Singer, 2015), an fmri compatible video task designed to capture both affective and cognitive empathy components in individual behavioural scores and associated neurophysiological activations.

3.2 Outlook and future studies

Acoustic correlates of social vocal control. Both emotional and social information is encoded in the acoustic vocal signal during speech (Hughes et al., 2014; Ko, Sadler, & Galinsky, 2014; Scherer, 1995). Particularly fundamental frequency (F0) and formant changes have been suggested to be correlates of socio-emotional information in the acoustic signal (see Pisanski, Cartei, et al., 2016 for review). But also glottal features and their perceptual correlates voice quality as well as temporal features of vocalizations have been related to production of socio-emotional attributes (Brück, Kreifelts, & Wildgruber, 2011; Liu & Xu, 2011; Xu, Kelly, & Smillie, 2013). How can these parameters be volitionally modulated? Vocal tract length can be exaggerated by lip protrusion, for instance, and by lowering the larynx. In fact, Lee et al. (2006) were able to show differential changes in vocal tract shape, length and variability depending on the emotional inflection of the produced speech. Although in a single subject, relating this methodology to neural activation patterns during socio-emotional voice modulation could offer the missing insights into socio-emotional vocal flexibility and its neural mechanisms.

The combination of acoustic vocal analyses and naïve ratings could help refine the signals central to effective social vocal communication, i.e. through the analysis of lens-models (e.g. Ko et al., 2014), which allow the acoustic parameters used by the speaker and listener for communication to be differentiated. In fact, Pardo et al. (2013) suggest to use naïve listener ratings as regressors on acoustic parameters, in contrast to traditional acoustic analysis. This is in line with previous work showing that listener ratings are more sensitive to the paralinguistic content of speech than acoustic parameters (Banse & Scherer, 1996) or uncorrelated to these (Hubbard & Trauner, 2007). Recent work has started to focus on the combination of listener ratings and acoustic analysis within the framework of reverse correlations (Ponsot et al., 2018). In addition to elucidating the effect of acoustic parameters in social nonverbal communication, a refinement of the contribution of particular acoustic dynamics can also inform mechanisms on the neurophysiological level of vocal control, for instance, by using multivariate analyses such as multi-voxel pattern analysis (MVPA) and representational similarity analysis (RSA; Kriegeskorte, Mur, & Bandettini, 2008). Both naïve rating data and acoustic data could be combined to refine specific areas involved in motor and auditory/somatosensory (feedforward and feedback) processing during social vocal control.

Vocal behaviour in dyadic interaction

Vocal flexibility can be exploited to meet a speaker's interests, be it to communicate effectively or to elicit a desirable response from a listener. The work presented in this thesis elucidates how social vocal control is achieved in isolation. But of course the voice is most prominently used in direct social interactions, and future investigations both on the behavioural as well as the neurophysiological level should take this into account (Redcay & Schilbach, 2019; Schilbach, 2016) by investigating voice change in life conversation.

In dyadic scenarios, speakers often converge to each other in their paralinguistic cues, for instance in their voice pitch and intensity (e.g. Babel, 2012; Gregory & Webster, 1996; Natale, 1975; Pardo, 2006). This type of vocal alignment has a social function, as it is positively associated with affiliation and liking (Adank, Stewart, Connell, & Wood, 2013; Lakin, Jefferis, Cheng, & Chartrand, 2003; Pardo, 2006; Pardo et al., 2012) and perceived quality of an interaction (Gregory et al., 1997; Gregory et al., 1993). This has been conceptualized in Communication Accommodation Theory (CAT; Giles et al., 1991). CAT suggests that vocal alignment is a dynamic process supporting social interaction, reflected in the synchronization of higher-order cognitive regions between interlocutors (Dikker, Silbert, Hasson, & Zevin, 2014; Hasson, Ghazanfar, Galantucci, Garrod, & Keysers, 2012; Silbert, Honey, Simony, Poeppel, & Hasson, 2014; Stephens, Silbert, & Hasson, 2010). It can be seen as a type of imitative behaviour, that occurs often in conversation, sometimes outside of conscious awareness (Garnier, Lamalle, & Sato, 2013; Pardo, 2013). However, vocal convergence is also subject to situational and social contexts (Pardo et al., 2010) and vocal divergence can be an equally effective sign as alignment can be (Healey, Purver, & Howes, 2014). The social desirability and status of the interlocutors is predictive of the direction of convergence or divergence (Natale, 1975), such that convergence occurs mostly in the direction of the speaker with higher social status (Gregory & Webster, 1996). Beyond signalling such social information, vocal alignment on the socio-emotional level can also increase emotional transfer between individuals by acting as affect induction signals, which evoke a concordant emotional state in an interlocutor (Neumann & Strack, 2000). Alignment can therefore serve to inform interlocutors of the other's emotions and mental state, and enhance bonding and understanding. Thus, similar to the notion of linguistic alignment facilitating linguistic speech comprehension (Garrod & Pickering, 2004), emotional alignment reflected in socio-emotional vocal patterns might be at the basis of successful social communication (Frith & Hasson, 2016).

Development of social vocal control

Speech is a foremost social behaviour and as such, how we express ourselves vocally develops in social interactions from early childhood onwards. There is evidence that we learn to express socially relevant information in our voice over ontogeny (Cartei, Cowles, Banerjee, & Reby, 2014; Redford & Gildersleeve-Neumann, 2009; for review see Locke, 2017), through imitation (Kuhl & Meltzoff, 1996; Sato et al., 2013) to approximate a vocal target through sensorimotor experience (Catmur, Walsh, & Heyes, 2009). Imitation in and of itself is a social behaviour highly associated with affiliation and bonding (Lakin et al., 2003), which engages neural reward processing areas (Kühn et al., 2010). Early sensorimotor experience might be acquired through imitative interactions with caregivers (Heyes, 2010; Paulus, 2014) and later on through as-if play, in which children often practice societal roles extensively. Experience has been shown to affect both linguistic abilities and nonverbal communication skills in children, depending on their interactions with caregivers or siblings, respectively (Havron et al., 2019; Hoff, 2006). Children learn to speak in voices in correspondence with stereotypical societal (Cartei et al., 2020) and gender roles (Cartei et al., 2014). On a neural level, vocal learning might be realized through the acquisition of sound and sensorimotor maps through practice, thus shaping motor plans for vocal expression through imitation (Tourville & Guenther, 2011). Associative sequence learning (ASL) theory suggests that imitation is automatic and thereby supports social learning particularly through the engagement of cortico-striate (Belyk, Pfordresher, Liotti, & Brown, 2016) or mirror neuron networks (Iacoboni et al., 2009), but that it can also be voluntarily controlled via top-down regulation, most likely reflected in inhibitory control from mentalizing processes in mPFC and TPJ (Wang & Hamilton, 2012).

Using the voice in social interactions: The Vocal Triflex

In summary, this work has focussed on voluntary modulation of the voice to express social information as one subcomponent of vocal flexibility. Based on the findings provided here, we suggest that vocal flexibility relies on the cooperation of perceptual and productive processes and we propose a triple-process model of vocal flexibility (Vocal Triflex) to inspire novel hypotheses. The model includes three main nodes which allow flexible vocal behaviour: perceptual, imitative and modulatory processes. The accurate perception of vocal

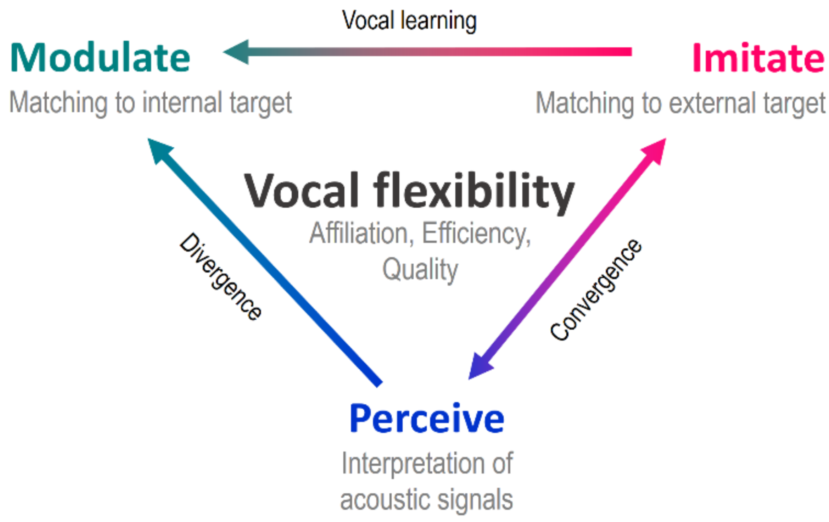


Figure 3. Illustration of a triple-process model of vocal flexibility. Perceptual, modulatory and imitative processes interact to allow for flexible vocal behaviour in social interactions. Vocal flexibility supports social signalling and communicative efficiency. While the perceptual node supports the interpretation of others voices, it also informs a speakers feedforward maps to diverge or converge to interlocutors. Converging to others' speech engages imitative processes, which might in turn inform the interpretation of others' speech signals. Through ontogeny imitation supports the establishment of internal speech sound maps that can be used independently to modulate vocal behaviour in response to the interlocutor. All three processes can be attended to automatically and volitionally.

cues from others informs a speakers' behaviour to converge or diverge from other speakers. While convergence is supported by vocal mimicry to an external target, divergence is supported by vocal modulation based on an internal target. Internal targets become independent of external scripts through vocal learning from imitation over ontogeny. Through these three processes, vocal flexibility itself contributes to signalling affiliative information and supports communication efficiency and quality of an interaction. The suggested mechanisms involved in vocal flexible behaviour are illustrated and summarized in figure 3.

How might these processes be reflected in the brain? Imitative processes might be supported by the engagement of common networks engaged both during listening to and producing a voice (Aziz-Zadeh et al., 2010; Garnier et al., 2013), whereas modulatory processes might be supported by mPFC and precuneus harbouring social knowledge and scripts, as well as areas involved in episodic memory processing in hippocampal regions. Both voice production components might engage feedforward and feedback regions involved in motor planning and control, sensorimotor regions, and areas engaged in social processing such as TPJ. Perceptual processes on the other hand, might engage auditory processing streams including STG, pSTG and dorsolateral prefrontal areas, as well as social brain networks in STS supporting intention inference and affective voice processing (Dricu & Frühholz, 2016). Given the putative role of brain networks involved in social processing supporting social vocal control, this model also allows for hypotheses regarding neural processing strategies related to social communicative impairments in clinical populations.

Clinical implications

The RDoC (Insel et al., 2010) aim to define systems for key domains of human behaviour, which can be analysed on different system levels (e.g. self-report, behaviour, neural circuits) and argue that social interactive deficits are a prevalent and pervasive symptom cluster within a realm of mental disorders (Cuthbert & Insel, 2013). The production of non-facial social communication is a subdomain of the system for social communication. One important aspect of social interactions is the exchange of information via the visual and auditory modularity as well as their successful integration. Whilst visual processing of socio-emotional information in faces, body posture and gestures has been studied thoroughly (e.g. Wang & Hamilton, 2012), the degree to which the voice contributes to supporting the successful interaction remains elusive. In fact, abnormal vocal behaviour has been discussed as disorder specific biomarker for acute phases of mental illness (for instance in schizophrenia or depression, reviewed by (Cannizzaro, Harel, Reilly, Chappell, & Snyder, 2004; Foltz, Rosenstein, & Elvevåg, 2016). Automated voice analysis is moving into the focus of research as a diagnostic tool for mental disorders (Cohen & Elvevåg, 2014).

Clinical populations with difficulties in perception of socio-emotional information from voices are, for example, people with autism spectrum disorder (ASD; Gebauer, Skewes, Hørlyck, & Vuust, 2014; Paul, Augustyn, Klin, & Volkmar, 2005) or psychopathy (Bagley, Abramowitz, & Kosson, 2009; Blair et al., 2002). ASD is characterized by pervasive social impairment, repetitive behaviours and restricted interests (American Psychiatric Association, 2013). Moreover, ASD has traditionally been associated with an atypical speech style (Asperger, 1991), which is reflected in multivariate variation across acoustic parameters and particularly in the spontaneous modulation of pitch (Fusaroli et al., 2017) also in subclinical populations (Shriberg et al., 2001). However, past research has focussed on spontaneous vocal modulation. The work reported in this thesis suggests that empathy skills might be supporting volitional social voice changes, which are specifically impaired in ASD (Baron-Cohen & Wheelwright, 2004), on the neural level. Future work is needed to see whether social vocal control can be specifically targeted and improved in therapeutic interventions (Dunn, Harris, & Dunn, 2017). In fact, nonverbal communications trainings already exist, which encompass, albeit indirectly, social vocal behaviour (Dunn et al., 2017; Kreifelts et al., 2013).

Social vocal behaviour is important for both patients and therapists. Most psychotherapeutic approaches are based on spoken communication between patient and therapist. Firstly, the vocal behaviour of the patient might be highly informative of the patient's current mental state, or experienced emotion (e.g. Laukka, 2007; Sauter, Eisner, Calder, & Scott, 2010; Scherer, 1972). In ASD, for instance, the magnitude of vocal rigidity, i.e. the lack of vocal alignment with the therapist has been shown to be associated with diagnostic ratings (Bone et al., 2014). However, strikingly little work has systematically investigated vocal behaviour of patients in therapeutic settings and vocal behaviour receives little attention during clinical training. This becomes even more relevant when considering that, secondly, the vocal behaviour of the therapist is of course, also heard by the patient. Signalling empathic responding to the patient and using vocal alignment to "pace" the patient might contribute to the quality of the therapeutic relationship and ultimately, might support treatment success (Bady, 1985). Thus, the work regarding mechanisms of volitional vocal behaviour draws attention to how the therapist can effectively control how he or she is heard by the patient. Lastly, during therapy, the patient also hears him- or herself speaking. Both what is said and how it is said reflects the acute psychological experience of the patient. Drawing attention to one's own vocal behaviour can therefore be understood as a metacognitive act supporting insight into the patient's own mental processes – an important contributor to behavioural change (Bertau, 2008; Frith, 2012; Neff & Dahm, 2015).

3.3 Summary and conclusions

In summary, volitional vocal control to express social information is an important route through which speakers can manage how they are heard by listeners. In the work presented in this thesis, evidence is provided that social vocal control is effective and recognizable, and speakers can evoke amplifications of social traits in their voice along the dimensions of the social trait space. Moreover, we show that social processing streams support the expression of social information in the voice on the neural level. Although dispositional empathy did not predict social vocal control ability, cognitive and affective components of empathy might contribute to differential task strategies on the neural level. Lastly, a social-opportunistic interpersonal style (Machiavellistic traits) was associated with better performance in favourable trait expression (sounding likeable). Together these findings highlight the importance of volitional social vocal behaviour for social interactions.

"Whether personality is expressed as adequately in the voice as in gesture or in carriage, we do not know. Perhaps it is even more adequately expressed in the voice than in these. In any event, it is clear that the nervous processes that control voice production must share in the individual traits of the nervous organization that condition the personality."

Edward Sapir, 1927, in *Speech as a Personality Trait*, p. 896

References

Ackermann, H., Hage, S. R., & Ziegler, W. (2014). Brain mechanisms of acoustic communication in humans and nonhuman primates: An evolutionary perspective. *Behavioral and Brain Sciences*, *72*(06), 1–84. https://doi.org/10.1017/S0140525X13003099

Adank, P., Stewart, A. J., Connell, L., & Wood, J. (2013). Accent imitation positively affects language attitudes. *Frontiers in Psychology*, *4*(280). https://doi.org/10.3389/fpsyg.2013.00280

Amenta, S., Noël, X., Verbanck, P., & Campanella, S. (2013). Decoding of Emotional Components in Complex Communicative Situations (Irony) and Its Relation to Empathic Abilities in Male Chronic Alcoholics: An Issue for Treatment. *Alcoholism: Clinical and Experimental Research*, *37*(2), 339–347. https://doi.org/10.1111/j.1530-0277.2012.01909.x

American Psychiatric Association. (2013). *Diagnostic and statistical manual of mental disorders*. (Author, Ed.) (5th ed.). Arlington.

Asperger, H. (1991). *"Autistic psychopathy" in childhood.* (U. Frith, Ed.), *Autism and Asperger syndrome.* New York, NY, US: Cambridge University Press. https://doi.org/10.1017/CBO9780511526770.002

Aziz-Zadeh, L., Sheng, T., & Gheytanchi, A. (2010). Common premotor regions for the perception and production of prosody and correlations with empathy and prosodic ability. *PLoS ONE*, *5*(1), e8759. https://doi.org/10.1371/journal.pone.0008759

Babel, M. (2012). Evidence for phonetic and social selectivity in spontaneous phonetic imitation. *Journal of Phonetics*, *40*(1), 177–189. https://doi.org/10.1016/j.wocn.2011.09.001

Bady, S. L. (1985). The voice as a curative factor in psychotherapy. *Psychoanalytic Review*, *72*(3), 479–490.

Bagley, A. D., Abramowitz, C. S., & Kosson, D. S. (2009). Vocal affect recognition and psychopathy: converging findings across traditional and cluster analytic approaches to assessing the construct. *Journal of Abnormal Psychology*, *118*(2), 388–398. https://doi.org/10.1037/a0015372

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology*, *70*(3), 614–636. https://doi.org/10.1037/0022-3514.70.3.614

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. https://doi.org/10.1023/B:JADD.0000022607.19833.00

Barrett, J., Pike, G. B., & Paus, T. (2004). The role of the anterior cingulate cortex in pitch variation during sad affect. *European Journal of Neuroscience*, *19*(2), 458–464. https://doi.org/10.1111/j.0953-816X.2003.03113.x

Baus, C., McAleer, P., Marcoux, K., Belin, P., & Costa, A. (2019). Forming social impressions from voices in native and foreign languages. *Scientific Reports*, *9*(1), 414.

https://doi.org/10.1038/s41598-018-36518-6

Belin, P., Boehme, B., & McAleer, P. (2017). The sound of trustworthiness: Acoustic-based modulation of perceived voice personality. *PLoS ONE*, *12*(10), e0185651. https://doi.org/10.1371/journal.pone.0185651

Belyk, M., & Brown, S. (2016). Pitch underlies activation of the vocal system during affective vocalization. *Social Cognitive and Affective Neuroscience*, *11*(7), 1078–1088. https://doi.org/10.1093/scan/nsv074

Belyk, M., Pfordresher, P. Q., Liotti, M., & Brown, S. (2016). The Neural Basis of Vocal Pitch Imitation in Humans. *Journal of Cognitive Neuroscience*, *28*(4), 621–635. https://doi.org/10.1162/jocn_a_00914

Berry, D. S. (1990). Vocal attractiveness and vocal babyishness: Effects on stranger, self, and friend impressions. *Journal of Nonverbal Behavior*, *14*(3), 141–153. https://doi.org/10.1007/BF00996223

Bertau, M. C. (2008). Voice: A pathway to consciousness as “social contact to oneself.” *Integrative Psychological and Behavioral Science*, *42*(1), 92–113. https://doi.org/10.1007/s12124-007-9041-8

Bird, G., & Viding, E. (2014). The self to other model of empathy: Providing a new framework for understanding empathy impairments in psychopathy, autism, and alexithymia. *Neuroscience and Biobehavioral Reviews*, *47*, 520–532. https://doi.org/10.1016/j.neubiorev.2014.09.021

Blair, R. J. R., Mitchell, D. G. V, Richell, R. A., Kelly, S., Leonard, A., Newman, C., & Scott, S. K. (2002). Turning a deaf ear to fear: impaired recognition of vocal affect in psychopathic individuals. *Journal of Abnormal Psychology*, *111*(4), 682–686.

Bone, D., Lee, C.-C., Black, M. P., Williams, M. E., Lee, S., Levitt, P., & Narayanan, S. (2014). The Psychologist as an Interlocutor in Autism Spectrum Disorder Assessment: Insights From a Study of Spontaneous Prosody. *Journal of Speech, Language, and Hearing Research*, *57*(4), 1162–1177. https://doi.org/10.1044/2014_jslhr-s-13-0062

Book, A., Methot, T., Gauthier, N., Hosker-Field, A., Forth, A., Quinsey, V., & Molnar, D. (2015). The mask of sanity revisited: Psychopathic traits and affective mimicry. *Evolutionary Psychological Science*, *1*(2), 91–102. https://doi.org/10.1007/s40806-015-0012-x

Brown, S., Cockett, P., & Yuan, Y. (2019). The neuroscience of Romeo and Juliet: An fMRI study of acting. *Royal Society Open Science*, *6*(3), 181908. https://doi.org/10.1098/rsos.181908

Brück, C., Kreifelts, B., & Wildgruber, D. (2011). Emotional voices in context: A neurobiological model of multimodal affective information processing. *Physics of Life Reviews*, *8*(4), 383–403. https://doi.org/10.1016/j.plrev.2011.10.002

Burnham, D., Kitamura, C., & Vollmer-Conna, U. (2002). What’s New, Pussycat? OnTalking to Babies and Animals. *Science*, *296*(5572), 1435. https://doi.org/10.1126/science.1069587

Bzdok, D., Langner, R., Schilbach, L., Engemann, D. A., Laird, A. R., Fox, P. T., & Eickhoff, S. B. (2013). Segregation of the human medial prefrontal cortex in social cognition. *Frontiers in Human Neuroscience*, *7*(232).

https://doi.org/10.3389/fnhum.2013.00232

Cabeza, R., Ciaramelli, E., & Moscovitch, M. (2012). Cognitive contributions of the ventral parietal cortex: An integrative theoretical account. *Trends in Cognitive Sciences*, *16*(6), 338–352. https://doi.org/10.1016/j.tics.2012.04.008

Calvo-Merino, B., Glaser, D. E., Grèzes, J., Passingham, R. E., & Haggard, P. (2005). Action observation and acquired motor skills: An fMRI study with expert dancers. *Cerebral Cortex*, *15*(8), 1243–1249. https://doi.org/10.1093/cercor/bhi007

Cannizzaro, M., Harel, B., Reilly, N., Chappell, P., & Snyder, P. J. (2004). Voice acoustical measurement of the severity of major depression. *Brain and Cognition*, *56*(1), 30–35. https://doi.org/10.1016/j.bandc.2004.05.003

Carey, D., & McGettigan, C. (2017). Magnetic resonance imaging of the brain and vocal tract: Applications to the study of speech production and language learning. *Neuropsychologia*, *98*, 201–211. https://doi.org/10.1016/j.neuropsychologia.2016.06.003

Cartei, V., Cowles, W., Banerjee, R., & Reby, D. (2014). Control of voice gender in pre-pubertal children. *British Journal of Developmental Psychology*, *32*(1), 100–106. https://doi.org/10.1111/bjdp.12027

Cartei, V., Oakhill, J., Garnham, A., Banerjee, R., & Reby, D. (2020). “This Is What a Mechanic Sounds Like”: Children’s Vocal Control Reveals Implicit Occupational Stereotypes. *Psychological Science*, *31*(8), 957–967. https://doi.org/10.1177/0956797620929297

Catmur, C., Walsh, V., & Heyes, C. (2009). Associative sequence learning: the role of experience in the development of imitation and the mirror system. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*(1528), 2369–2380. https://doi.org/10.1098/rstb.2009.0048

Chebat, J.-C., El Hedhli, K., Gélinas-Chebat, C., & Boivin, R. (2007). Voice and persuasion in a banking telemarketing context. *Perceptual and Motor Skills*, *104*(2), 419–437. https://doi.org/10.2466/pms.104.2.419-437

Christie, R., & Geis, F. L. (1970). *Studies in Machiavellianism*. *Studies in Machiavellianism*. New York: Academic Press. https://doi.org/10.1016/c2013-0-10497-7

Cohen, A. S., & Elvevåg, B. (2014). Automated computerized analysis of speech in psychiatric disorders. *Current Opinion in Psychiatry*, *27*(3), 203–209. https://doi.org/10.1097/YCO.0000000000000056

Corral, S., & Calvete, E. (2000). Machiavellianism: Dimensionality of the Mach IV and its Relation to Self-Monitoring in a Spanish Sample. *Spanish Journal of Psychology*, *3*(1), 3–13. https://doi.org/10.1017/S1138741600005497

Cuff, B. M. P., Brown, S. J., Taylor, L., & Howat, D. J. (2014). Empathy: A review of the concept. *Emotion Review*, *8*(2), 144–153. https://doi.org/10.1177/1754073914558466

Cuthbert, B. N., & Insel, T. R. (2013). Toward the future of psychiatric diagnosis: The seven pillars of RDoC. *BMC Medicine*, *11*(1), 126. https://doi.org/10.1186/1741-7015-11-126

Dahl, A., Sherlock, B. R., Campos, J. J., & Theunissen, F. E. (2014). Mothers’ tone of voice depends on the nature of infants’ transgressions. *Emotion*, *14*(4), 651–665.

https://doi.org/10.1037/a0036608

de Guzman, M., Bird, G., Banissy, M. J., & Catmur, C. (2016). Self–other control processes in social cognition: From imitation to empathy. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150079. https://doi.org/10.1098/rstb.2015.0079

Dichter, B. K., Breshears, J. D., Leonard, M. K., & Chang, E. F. (2018). The Control of Vocal Pitch in Human Laryngeal Motor Cortex. *Cell*, *174*(1), 21-31.e9. https://doi.org/10.1016/j.cell.2018.05.016

Dikker, S., Silbert, L. J., Hasson, U., & Zevin, J. D. (2014). On the same wavelength: predictable language enhances speaker-listener brain-to-brain synchrony in posterior superior temporal gyrus. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *34*(18), 6267–6272. https://doi.org/10.1523/JNEUROSCI.3796-13.2014

Dogil, G., Ackermann, H., Grodd, W., Haider, H., Kamp, H., Mayer, J., … Wildgruber, D. (2002). The speaking brain: A tutorial introduction to fMRI experiments in the production of speech, prosody and syntax. *Journal of Neurolinguistics*, *15*(1), 59–90. https://doi.org/10.1016/S0911-6044(00)00021-X

Dricu, M., & Frühholz, S. (2016). Perceiving emotional expressions in others: Activation likelihood estimation meta-analyses of explicit evaluation, passive perception and incidental perception of emotions. *Neuroscience and Biobehavioral Reviews*, *71*, 810–828. https://doi.org/10.1016/j.neubiorev.2016.10.020

Dunn, M. A., Harris, L., & Dunn, J. (2017). *Prosody Intervention for High-Functioning Adolescents and Adults with Autism Spectrum Disorder : Enhancing Communication and Social Engagement Through Voice, Rhythm, and Pitch*. London and Philadelphia: Jessica Kingsley Publishers.

Elbert, S. P., & Dijkstra, A. (2014). An experimental test of the relationship between voice intonation and persuasion in the domain of health. *Psychology and Health*, *29*(9), 1014–1031. https://doi.org/10.1080/08870446.2014.903482

Eriksson, A., Llamas, C., & Watt, D. (2010). The Disguised Voice : Imitating Accents or Speech Styles and Impersonating Individuals. In *Language and Identities* (pp. 86–96). Edinburgh: Edinburgh University Press.

Fant, G. (1960). *Acoustic Theory of Speech Production*. The Hague: Mouton & Co.

Farley, S. D., Hughes, S. M., & LaFayette, J. N. (2013). People Will Know We Are in Love: Evidence of Differences Between Vocal Samples Directed Toward Lovers and Friends. *Journal of Nonverbal Behavior*, *37*(3), 123–138. https://doi.org/10.1007/s10919-013-0151-3

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. https://doi.org/10.1016/j.anbehav.2004.06.012

Fiske, S. T., Cuddy, A. J. C., Glick, P., & Xu, J. (2002). A model of (often mixed) stereotype content: Competence and warmth respectively follow from perceived status and competition. *Journal of Personality and Social Psychology*, *82*(6), 878–902.

https://doi.org/10.1037/0022-3514.82.6.878

Foltz, P. W., Rosenstein, M., & Elvevåg, B. (2016). Detecting clinically significant events through automated language analysis: Quo imus? *NPJ Schizophrenia*, *2*, 15054. https://doi.org/10.1038/npjschz.2015.54

Frith, C. D. (2012). The role of metacognition in human social interactions. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1599), 2213–2223. https://doi.org/10.1098/rstb.2012.0123

Frith, C. D., & Hasson, U. (2016). Mirroring and beyond: coupled dynamics as a generalized framework for modelling social interactions. *Royal Society Philosphical Transactions B*, *371*(1693), 20150366. https://doi.org/10.1098/rstb.2015.0366

Frith, U., & Frith, C. D. (2003). Development and neurophysiology of mentalizing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *358*(1431), 459–473. https://doi.org/10.1098/rstb.2002.1218

Frühholz, S., Klaas, H. S., Patel, S., & Grandjean, D. (2015). Talking in fury: The cortico-subcortical network underlying angry vocalizations. *Cerebral Cortex*, *25*(9), 2752–2762. https://doi.org/10.1093/cercor/bhu074

Fusaroli, R., Lambrechts, A., Bang, D., Bowler, D. M., & Gaigg, S. B. (2017). Is voice a marker for Autism spectrum disorder? A systematic review and meta-analysis. *Autism Research*, *10*(3), 384–407. https://doi.org/10.1002/aur.1678

Gangestad, S. W., & Snyder, M. (2000). Self-monitoring: Appraisal and reappraisal. *Psychological Bulletin*, *126*(4), 530–555. https://doi.org/10.1037/0033-2909.126.4.530

Garnier, M., & Henrich, N. (2014). Speaking in noise: How does the Lombard effect improve acoustic contrasts between speech and ambient noise? *Computer Speech & Language*, *28*(2), 580–597. https://doi.org/10.1016/j.csl.2013.07.005

Garnier, M., Lamalle, L., & Sato, M. (2013). Neural correlates of phonetic convergence and speech imitation. *Frontiers in Psychology*, *4*(600). https://doi.org/10.3389/fpsyg.2013.00600

Gebauer, L., Skewes, J., Hørlyck, L., & Vuust, P. (2014). Atypical perception of affective prosody in Autism Spectrum Disorder. *NeuroImage. Clinical*, *6*, 370–378. https://doi.org/10.1016/j.nicl.2014.08.025

Giles, H., Coupland, N., & Coupland, I. (1991). 1. Accommodation theory: Communication, context, and consequence. In H. Giles, N. Coupland, & I. Coupland (Eds.), *Contexts of Accommodation: Developments in Applied Sociolinguistics* (pp. 1–68). New York, NY: Cambridge University Press.

Gregory, S. W., Dagan, K. a., & Webster, S. (1997). Evaluating the relation of vocal accommodation in conversation partners' fundamental requencies to perception of communication quality. *Journal of Nonverbal Behavior*, *21*(1), 23–43. https://doi.org/10.1023/A:1024995717773

Gregory, S. W., & Webster, S. (1996). A nonverbal signal in voices of interview partners effectively predicts communication accommodation and social status perceptions. *Journal of Personality and Social Psychology*, *70*(6), 1231–1240. https://doi.org/10.1037/0022-3514.70.6.1231

Gregory, S., Webster, S., & Huang, G. (1993). Voice pitch and amplitude convergence as a metric of quality in dyadic interviews. *Language and Communication*, *13*(3), 195–217. https://doi.org/10.1016/0271-5309(93)90026-J

Grice, H. P. (1957). Meaning. *The Philosophical Review*, *66*(3), 377–388.

Grice, H. P. (1975). Logic and conversation. In P. Cole & J. Morgan (Eds.), *Syntax and semantics* (pp. 41–58). New York: Academic Press.

Grieve, R. (2011). Mirror Mirror: The role of self-monitoring and sincerity in emotional manipulation. *Personality and Individual Differences*, *51*(8), 981–985. https://doi.org/10.1016/j.paid.2011.08.004

Hammerschmidt, K., & Jürgens, U. (2007). Acoustical Correlates of Affective Prosody. *Journal of Voice*, *21*(5), 531–540. https://doi.org/10.1016/j.jvoice.2006.03.002

Hasson, U., Ghazanfar, A. A., Galantucci, B., Garrod, S., & Keysers, C. (2012). Brain-to-brain coupling: A mechanism for creating and sharing a social world. *Trends in Cognitive Sciences*, *16*(2), 114–121. https://doi.org/10.1016/j.tics.2011.12.007

Havron, N., Ramus, F., Heude, B., Forhan, A., Cristia, A., Peyre, H., … Thiebaugeorges, O. (2019). The Effect of Older Siblings on Language Development as a Function of Age Difference and Sex. *Psychological Science*, *30*(9), 1333–1343. https://doi.org/10.1177/0956797619861436

Healey, P. G. T., Purver, M., & Howes, C. (2014). Divergence in dialogue. *PloS One*, *9*(2), e98598. https://doi.org/10.1371/journal.pone.0098598

Hecht, M. A., & LaFrance, M. (1995). How (Fast) Can I Help You? Tone of Voice and Telephone Operator Efficiency in Interactions. *Journal of Applied Social Psychology*, *25*(23), 2086–2098. https://doi.org/10.1111/j.1559-1816.1995.tb02389.x

Hellbernd, N., & Sammler, D. (2016). Prosody conveys speaker's intentions: Acoustic cues for speech act perception. *Journal of Memory and Language*, *88*, 70–86. https://doi.org/10.1016/j.jml.2016.01.001

Hellbernd, N., & Sammler, D. (2018). Neural bases of social communicative intentions in speech. *Social Cognitive and Affective Neuroscience*, *13*(6), 604–615. https://doi.org/10.1093/scan/nsy034

Heyes, C. (2010). Where do mirror neurons come from? *Neuroscience and Biobehavioral Reviews*, *34*(4), 575–583. https://doi.org/10.1016/j.neubiorev.2009.11.007

Hirson, A., & Duckworth, M. (1993). Glottal fry and voice disguise: a case study in forensic phonetics. *Journal of Biomedical Engineering*, *15*(3), 193–200. https://doi.org/10.1016/0141-5425(93)90115-F

Hoff, E. (2006). How social contexts support and shape language development. *Developmental Review*, *26*(1), 55–88. https://doi.org/10.1016/j.dr.2005.11.002

Hubbard, K., & Trauner, D. A. (2007). Intonation and emotion in autistic spectrum disorders. *Journal of Psycholinguistic Research*, *36*(2), 159–173. https://doi.org/10.1007/s10936-006-9037-4

Hughes, S. M., & Miller, N. E. (2016). What sounds beautiful looks beautiful stereotype. *Journal of Social and Personal Relationships*, *33*(7), 984–996.

https://doi.org/10.1177/0265407515612445

Hughes, S. M., Mogilski, J. K., & Harrison, M. A. (2014). The Perception and Parameters of Intentional Voice Manipulation. *Journal of Nonverbal Behavior*, *38*(1), 107–127. https://doi.org/10.1007/s10919-013-0163-z

Iacoboni, M., Iacoboni, M., Woods, R. P., Brass, M., Bekkering, H., Mazziotta, J. C., & Rizzolatti, G. (2009). Cortical Mechanisms of Human Imitation. *Science*, *2526*(1999). https://doi.org/10.1126/science.286.5449.2526

Insel, T., Cuthbert, B., Garvey, M., Heinssen, R., Pine, D. S., Quinn, K., … Wang, P. (2010). Research Domain Criteria (RDoC): Toward a new classification framework for research on mental disorders. *American Journal of Psychiatry*, *167*(7), 748–751. https://doi.org/10.1176/appi.ajp.2010.09091379

Jacob, H., Kreifelts, B., Nizielski, S., Schütz, A., & Wildgruber, D. (2016). Effects of emotional intelligence on the impression of irony created by the mismatch between verbal and nonverbal cues. *PLoS ONE*, *11*(10). https://doi.org/10.1371/journal.pone.0163211

Jansen, W., Gregory, M. L., & Brenier, J. M. (2001). Prosodic correlates of directly reported speech: Evidence from conversational speech. In *Proceedings of the ISCA Workshop on Prosody in Speech Recognition and Understanding* (pp. 77–80).

Jiang, X., Sanford, R., & Pell, M. D. (2017). Neural systems for Evaluating speaker (Un)believability. *Human Brain Mapping*, *38*(7), 3732–3749. https://doi.org/10.1002/hbm.23630

Jonason, P. K., & Webster, G. D. (2012). A protean approach to social influence: Dark Triad personalities and social influence tactics. *Personality and Individual Differences*, *52*(4), 521–526. https://doi.org/10.1016/j.paid.2011.11.023

Jones, D., & Paulhus, D. (2009). Machiavellianism. Guilford Press.

Jürgens, U. (2009). The Neural Control of Vocalization in Mammals: A Review. *Journal of Voice*, *23*(1), 1–10. https://doi.org/10.1016/j.jvoice.2007.07.005

Jürgens, U., & Pratt, R. (1979). Role of the periaqueductal grey in vocal expression of emotion. *Brain Research*, *167*(2), 367–378. https://doi.org/10.1016/0006-8993(79)90830-8

Jürgens, U., & von Cramon, D. (1982). On the role of the anterior cingulate cortex in phonation: A case report. *Brain and Language*, *15*(2), 234–248. https://doi.org/10.1016/0093-934X(82)90058-X

Jürgens, Uwe. (2002). Neural pathways underlying vocal control. *Neuroscience and Biobehavioral Reviews*, *26*(2), 235–258. https://doi.org/10.1016/S0149-7634(01)00068-9

Jürgens, Uwe, & Zwirner, P. (1996). The role of the periaqueductal grey in limbic and neocortical vocal fold control. *NeuroReport*, *7*(18), 2921–2923. https://doi.org/10.1097/00001756-199611250-00023

Kanske, P., Böckler, A., Trautwein, F. M., & Singer, T. (2015). Dissecting the social brain: Introducing the EmpaToM to reveal distinct neural networks and brain-behavior relations for empathy and Theory of Mind. *NeuroImage*, *122*, 6–19.

https://doi.org/10.1016/j.neuroimage.2015.07.082

Karpf, A. (2007). *The Human Voice: The Story of a Remarkable Talent*. London: Bloomsbury.

Kemper, S., Finter-Urczyk, A., Ferrell, P., Harden, T., & Billington, C. (1998). Using elderspeak with older adults. *Discourse Processes*, *25*(1), 55–73. https://doi.org/10.1080/01638539809545020

Kiese-Himmel, C., Himmel, W., & Scherer, M. (2012). Is a change in vocal loudness a first step towards becoming a medical doctor? *Swiss Medical Weekly*, *142*(w13534). https://doi.org/10.4414/smw.2012.13534

Kitamura, C., & Burnham, D. (2003). Pitch and communicative intent in mother's speech: Adjustments for age and sex in the first year. *Infancy*, *4*(1), 85–110. https://doi.org/10.1207/S15327078IN0401_5

Klasen, M., von Marschall, C., Isman, G., Zvyagintsev, M., Gur, R. C., & Mathiak, K. (2018). Prosody production networks are modulated by sensory cues and social context. *Social Cognitive and Affective Neuroscience*, *13*(4), 418–429. https://doi.org/10.1093/scan/nsy015

Ko, S. J., Sadler, M. S., & Galinsky, A. D. (2014). The Sound of Power: Conveying and Detecting Hierarchical Rank Through Voice. *Psychological Science*, *26*(1), 3–14. https://doi.org/10.1177/0956797614553009

Krauss, R. M., Freyberg, R., & Morsella, E. (2002). Inferring speakers' physical attributes from their voices. *Journal of Experimental Social Psychology*, *38*(6), 618–625. https://doi.org/10.1016/S0022-1031(02)00510-3

Krauss, R. M., & Pardo, J. S. (2006). Bridging Social Psychology. In P. A. M. Van Lange (Ed.), *Bridging Social Psychology: Benefits of Transdisciplinary Approaches* (pp. 273–278). New York: Psychology Press. https://doi.org/10.4324/9781410616982

Kreifelts, B., Jacob, H., Carolin, B., Erb, M., Ethofer, T., & Wildgruber, D. (2013). Non-verbal emotion communication training induces specific changes in brain function and structure. *Frontiers in Human Neuroscience*, *7*(648). https://doi.org/10.3389/fnhum.2013.00648

Kriegeskorte, N., Mur, M., & Bandettini, P. (2008). Representational similarity analysis - connecting the branches of systems neuroscience. *Frontiers in Systems Neuroscience*, *2*, 4. https://doi.org/10.3389/neuro.06.004.2008

Krishnan, S., Lima, C. F., Evans, S., Chen, S., Guldner, S., Yeff, H., … Scott, S. K. (2018). Beatboxers and Guitarists Engage Sensorimotor Regions Selectively When Listening to the Instruments They can Play. *Cerebral Cortex*, *28*(11), 4063–4079. https://doi.org/10.1093/cercor/bhy208

Krueger, F., Barbey, A. K., & Grafman, J. (2009). The medial prefrontal cortex mediates social event knowledge. *Trends in Cognitive Sciences*, *13*(3), 103–109. https://doi.org/10.1016/j.tics.2008.12.005

Kuhl, P. K., & Meltzoff, A. N. (1996). Infant vocalizations in response to speech: Vocal imitation and developmental change. *The Journal of the Acoustical Society of America*, *100*(4), 2425–2438. https://doi.org/10.1121/1.417951

Kuhn, L. K., Wydell, T., Lavan, N., McGettigan, C., & Garrido, L. (2017). Similar representations of emotions across faces and voices. *Emotion*, *17*(6), 912–937. https://doi.org/10.1037/emo0000282

Kühn, S., Müller, B. C. N., van Baaren, R. B., Wietzker, A., Dijksterhuis, A., & Brass, M. (2010). Why do I like you when you behave like me? Neural mechanisms mediating positive consequences of observing someone being imitated. *Social Neuroscience*, *5*(4), 384–392. https://doi.org/10.1080/17470911003633750

Lakin, J. L., Jefferis, V. E., Cheng, C. M., & Chartrand, T. L. (2003). The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of Nonverbal Behavior*, *27*(3), 145–162. https://doi.org/10.1023/A:1025389814290

Laukka, P. (2007). Research on vocal expression of emotion: State of the art and future directions. In K. Izdebski (Ed.), *Emotions in the Human Voice: Foundations* (pp. 153–170). San Diego: Plural Publishing.

Laukka, Petri, Åhs, F., Furmark, T., & Fredrikson, M. (2011). Neurofunctional correlates of expressed vocal affect in social phobia. *Cognitive, Affective and Behavioral Neuroscience*, *11*(3), 413–425. https://doi.org/10.3758/s13415-011-0032-3

Lavan, N., Burston, L. F. K., Ladwa, P., Merriman, S. E., Knight, S., & McGettigan, C. (2019). Breaking voice identity perception: Expressive voices are more confusable for listeners. *Quarterly Journal of Experimental Psychology*, *72*(9), 2240–2248. https://doi.org/10.1177/1747021819836890

Lavan, N., Burton, A. M., Scott, S. K., & McGettigan, C. (2019). Flexible voices: Identity perception from variable vocal signals. *Psychonomic Bulletin and Review*, *26*(1), 90–102. https://doi.org/10.3758/s13423-018-1497-7

Lavan, N., Mileva, M., & McGettigan, C. (2021). How does familiarity with a voice affect trait judgements? *British Journal of Psychology*. https://doi.org/10.1111/bjop.12454

Lee, S., Bresch, E., Adams, J., Kazemzadeh, A., & Narayanan, S. (2006). A study of emotional speech articulation using a fast magnetic resonance imaging technique. In *INTERSPEECH 2006 and 9th International Conference on Spoken Language Processing, INTERSPEECH 2006 - ICSLP* (Vol. 5, pp. 2234–2237).

Lieberman, M. D., Straccia, M. A., Meyer, M. L., Du, M., & Tan, K. M. (2019). Social, self, (situational), and affective processes in medial prefrontal cortex (MPFC): Causal, multivariate, and reverse inference evidence. *Neuroscience and Biobehavioral Reviews*, *99*, 311–328. https://doi.org/10.1016/j.neubiorev.2018.12.021

Liu, X., & Xu, Y. (2011). What makes a female voice attractive? In *ICPhS* (pp. 17–21).

Locke, J. L. (2017). Emancipation of the voice: Vocal complexity as a fitness indicator. *Psychonomic Bulletin and Review*, *24*(1), 232–237. https://doi.org/10.3758/s13423-016-1105-7

Louth, S. M., Williamson, S., Alpert, M., Pouget, E. R., & Hare, R. D. (1998). Acoustic distinctions in the speech of male psychopaths. *Journal of Psycholinguistic Research*, *27*(3), 375–384. https://doi.org/10.1023/A:1023207821867

Ma, N., Vandekerckhove, M., Van Hoeck, N., & Van Overwalle, F. (2012). Distinct recruitment of temporo-parietal junction and medial prefrontal cortex in behavior understanding and trait identification. *Social Neuroscience*, *7*(6), 591–605. https://doi.org/10.1080/17470919.2012.686925

Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *PLoS ONE*, *13*(10). https://doi.org/10.1371/journal.pone.0204991

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say "hello"? Personality impressions from brief novel voices. *PLoS ONE*, *9*(3), e90779. https://doi.org/10.1371/journal.pone.0090779

McGettigan, C., Eisner, F., Agnew, Z. K., Manly, T., Wisbey, D., & Scott, S. K. (2013). T'ain't what you say, it's the way that you say it-left insula and inferior frontal cortex work in interaction with superior temporal regions to control the performance of vocal impersonations. *Journal of Cognitive Neuroscience*, *25*(11), 1875–1886. https://doi.org/10.1162/jocn_a_00427

Mileva, M., Tompkinson, J., Watt, D., & Burton, A. M. (2018). Audiovisual integration in social evaluation. *Journal of Experimental Psychology: Human Perception and Performance*, *44*(1), 128–138. https://doi.org/10.1037/xhp0000439

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. https://doi.org/10.1016/j.evolhumbehav.2016.10.010

Montepare, J. M., & Zebrowitz-McArthur, L. (1987). Perceptions of adults with childlike voices in two cultures. *Journal of Experimental Social Psychology*, *23*(4), 331–349. https://doi.org/10.1016/0022-1031(87)90045-X

Morningstar, M., Dirks, M. A., & Huang, S. (2017). Vocal Cues Underlying Youth and Adult Portrayals of Socio-emotional Expressions. *Journal of Nonverbal Behavior*, *41*(2), 155–183. https://doi.org/10.1007/s10919-017-0250-7

Muir, K., Joinson, A., Cotterill, R., & Dewdney, N. (2016). Characterizing the Linguistic Chameleon: Personal and Social Correlates of Linguistic Style Accommodation. *Human Communication Research*, *42*(3), 462–484. https://doi.org/10.1111/hcre.12083

Natale, M. (1975). Convergence of mean vocal intensity in dyadic communication as a function of social desirability. *Journal of Personality and Social Psychology*, *32*(5), 790–804. https://doi.org/10.1037/0022-3514.32.5.790

Neff, K. D., & Dahm, K. A. (2015). Self-compassion: What it is, what it does, and how it relates to mindfulness. In *Handbook of Mindfulness and Self-Regulation* (pp. 121–140). Springer New York. https://doi.org/10.1007/978-1-4939-2263-5_10

Neumann, R., & Strack, F. (2000). "Mood contagion": The automatic transfer of mood between persons. *Journal of Personality and Social Psychology*, *79*(2), 211–223. https://doi.org/10.1037/0022-3514.79.2.211

Neves, L., Cordeiro, C., Scott, S. K., Castro, S. L., & Lima, C. F. (2018). High emotional contagion and empathy are associated with enhanced detection of emotional authenticity in laughter. *Quarterly Journal of Experimental Psychology*, *71*(11), 2355–2363. https://doi.org/10.1177/1747021817741800

O'Connor, J. J. M., & Barclay, P. (2017). The influence of voice pitch on perceptions of trustworthiness across social contexts. *Evolution and Human Behavior*, *38*(4), 506–512.

https://doi.org/10.1016/j.evolhumbehav.2017.03.001

Olivola, C. Y., & Todorov, A. (2010). Elected in 100 milliseconds: Appearance-based trait inferences and voting. *Journal of Nonverbal Behavior*, *34*(2), 83–110. https://doi.org/10.1007/s10919-009-0082-1

Oosterhof, N. N., & Todorov, A. (2008). The functional basis of face evaluation. *Proceedings of the National Academy of Sciences of the United States of America*, *105*(32), 11087–11092. https://doi.org/10.1073/pnas.0805664105

Owren, M. J., Amoss, R. T., & Rendall, D. (2011). Two organizing principles of vocal production: Implications for nonhuman and human primates. *American Journal of Primatology*, *73*(6), 530–544. https://doi.org/10.1002/ajp.20913

Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, *119*(4), 2382–2393. https://doi.org/10.1121/1.2178720

Pardo, J. S. (2013). Measuring phonetic convergence in speech production. *Frontiers in Psychology*, *4*(559). https://doi.org/10.3389/fpsyg.2013.00559

Pardo, J. S., Gibbons, R., Suppes, A., & Krauss, R. M. (2012). Phonetic convergence in college roommates. *Journal of Phonetics*, *40*(1), 190–197. https://doi.org/10.1016/j.wocn.2011.10.001

Pardo, J. S., Jay, I. C., & Krauss, R. M. (2010). Conversational role influences speech imitation. *Attention, Perception, and Psychophysics*, *72*(8), 2254–2264. https://doi.org/10.3758/BF03196699

Paul, R., Augustyn, A., Klin, A., & Volkmar, F. R. (2005). Perception and production of prosody by speakers with autism spectrum disorders. *Journal of Autism and Developmental Disorders*, *35*(2), 205–220. https://doi.org/10.1007/s10803-004-1999-1

Paulhus, D. L., & Williams, K. M. (2002). The Dark Triad of personality: Narcissism, Machiavellianism, and psychopathy. *Journal of Research in Personality*, *36*(6), 556–563. https://doi.org/10.1016/S0092-6566(02)00505-6

Paulmann, S., Furnes, D., Bøkenes, A. M., & Cozzolino, P. J. (2016). How psychological stress affects emotional prosody. *PLoS ONE*, *11*(11), e0165022. https://doi.org/10.1371/journal.pone.0165022

Paulus, M. (2014). How and why do infants imitate? An ideomotor approach to social and imitative learning in infancy (and beyond). *Psychonomic Bulletin and Review*, *21*(5), 1139–1156. https://doi.org/10.3758/s13423-014-0598-1

Pavela Banai, I., Banai, B., & Bovan, K. (2017). Vocal characteristics of presidential candidates can predict the outcome of actual elections. *Evolution and Human Behavior*, *38*(3), 309–314. https://doi.org/10.1016/j.evolhumbehav.2016.10.012

Pichon, S., & Kell, C. A. (2013). Affective and sensorimotor components of emotional prosody generation. *The Journal of Neuroscience*, *33*(4), 1640–1650. https://doi.org/10.1523/JNEUROSCI.3530-12.2013

Pisanski, K., Cartei, V., McGettigan, C., Raine, J., & Reby, D. (2016). Voice Modulation: A Window into the Origins of Human Vocal Control? *Trends in Cognitive Sciences*, *20*(4), 304–318. https://doi.org/10.1016/j.tics.2016.01.002

Pisanski, K., Nowak, J., & Sorokowski, P. (2016). Individual differences in cortisol stress response predict increases in voice pitch during exam stress. *Physiology and Behavior*, *163*, 234–238. https://doi.org/10.1016/j.physbeh.2016.05.018

Pisanski, K., Oleszkiewicz, A., Plachetka, J., Gmiterek, M., & Reby, D. (2018). Voice pitch modulation in human mate choice. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1893), 20181634. https://doi.org/10.1098/rspb.2018.1634

Ponsot, E., Burred, J. J., Belin, P., & Aucouturier, J. J. (2018). Cracking the social code of speech prosody using reverse correlation. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(15), 3972–3977. https://doi.org/10.1073/pnas.1716090115

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. https://doi.org/10.1016/j.evolhumbehav.2005.11.003

Raphael, B. N., & Scherer, R. C. (1987). Voice modifications of stage actors: Acoustic analyses. *Journal of Voice*, *1*(1), 83–87. https://doi.org/10.1016/S0892-1997(87)80029-2

Redcay, E., & Schilbach, L. (2019). Using second-person neuroscience to elucidate the mechanisms of social interaction. *Nature Reviews Neuroscience*, *20*(8), 495–505. https://doi.org/10.1038/s41583-019-0179-4

Redford, M. A., & Gildersleeve-Neumann, C. E. (2009). The development of distinct speaking styles in preschool children. *Journal of Speech, Language, and Hearing Research*, *52*(6), 1434–1448. https://doi.org/10.1044/1092-4388(2009/07-0223)

Reniers, R. L. E. P., Corcoran, R., Drake, R., Shryane, N. M., & Völlm, B. A. (2011). The QCAE: A questionnaire of cognitive and affective empathy. *Journal of Personality Assessment*, *93*(1), 84–95. https://doi.org/10.1080/00223891.2010.528484

Sato, M., Grabski, K., Garnier, M., Granjon, L., Schwartz, J. L., & Nguyen, N. (2013). Converging toward a common speech code: Imitative and perceptuo-motor recalibration processes in speech production. *Frontiers in Psychology*, *4*(422). https://doi.org/10.3389/fpsyg.2013.00422

Sauter, D. A., Eisner, F., Calder, A. J., & Scott, S. K. (2010). Perceptual cues in nonverbal vocal expressions of emotion. *Quarterly Journal of Experimental Psychology*, *63*(11), 2251–2272. https://doi.org/10.1080/17470211003721642

Scherer, K. R. (1972). Judging personality from voice: A cross-cultural approach to an old issue in interpersonal perception. *Journal of Personality*, *40*(2), 191–210. https://doi.org/10.1111/j.1467-6494.1972.tb00998.x

Scherer, K. R. (1978). Personality inference from voice quality: The loud voice of extroversion. *European Journal of Social Psychology*, *8*(4), 467–487. https://doi.org/10.1002/ejsp.2420080405

Scherer, K. R. (1995). Expression of emotion in voice and music. *Journal of Voice*, *9*(3), 235–248. https://doi.org/10.1016/S0892-1997(05)80231-0

Schilbach, L. (2016). Towards a second-person neuropsychiatry. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1686), 20150081. https://doi.org/10.1098/rstb.2015.0081

Schroeder, J., & Epley, N. (2015). The Sound of Intellect: Speech Reveals a Thoughtful Mind, Increasing a Job Candidate's Appeal. *Psychological Science*, *26*(6), 877–891. https://doi.org/10.1177/0956797615572906

Schurz, M., Radua, J., Aichhorn, M., Richlan, F., & Perner, J. (2014). Fractionating theory of mind: A meta-analysis of functional brain imaging studies. *Neuroscience and Biobehavioral Reviews*, *42*, 9–34. https://doi.org/10.1016/j.neubiorev.2014.01.009

Scott, S., & McGettigan, C. (2016). The voice: From identity to interactions. In *APA handbook of nonverbal communication.* (pp. 289–305). Washington: American Psychological Association. https://doi.org/10.1037/14669-011

Shriberg, L. D., Paul, R., McSweeny, J. L., Klin, A., Cohen, D. J., & Volkmar, F. R. (2001). Speech and prosody characteristics of adolescents and adults with high-functioning autism and Asperger syndrome. *Journal of Speech, Language, and Hearing Research*, *44*(5), 1097–1115. https://doi.org/10.1044/1092-4388(2001/087)

Silbert, L. J., Honey, C. J., Simony, E., Poeppel, D., & Hasson, U. (2014). Coupled neural systems underlie the production and comprehension of naturalistic narrative speech. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(43), E4687-96. https://doi.org/10.1073/pnas.1323812111

Simonyan, K., & Horwitz, B. (2011). Laryngeal motor cortex and control of speech in humans. *Neuroscientist*, *17*(2), 197–208. https://doi.org/10.1177/1073858410386727

Skoog Waller, S., & Eriksson, M. (2016). Vocal age disguise: The role of fundamental frequency and speech rate and its perceived effects. *Frontiers in Psychology*, *7*(1814). https://doi.org/10.3389/fpsyg.2016.01814

Sorokowski, P., Puts, D., Johnson, J., Żółkiewicz, O., Oleszkiewicz, A., Sorokowska, A., … Pisanski, K. (2019). Voice of Authority: Professionals Lower Their Vocal Frequencies When Giving Expert Advice. *Journal of Nonverbal Behavior*, *43*(2), 257–269. https://doi.org/10.1007/s10919-019-00307-0

Sperber, D. (2000). Metarepresentations in an evolutionary perspective. In D. Sperber (Ed.), *Metarepresentations: A multidisciplinary perspective* (pp. 117–137). Oxford Univ. Press.

Stephens, G. J., Silbert, L. J., & Hasson, U. (2010). Speaker-listener neural coupling underlies successful communication. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(32), 14425–14430. https://doi.org/10.1073/pnas.1008662107

Streeter, L. A., Macdonald, N. H., Apple, W., Krauss, R. M., & Galotti, K. M. (1983). Acoustic and Perceptual Indicators of Emotional Stress. *Journal of Acoustical Society of America*, *73*(4), 1354–1360. https://doi.org/10.1121/1.389239

Tamir, D. I., & Thornton, M. A. (2018). Modeling the Predictive Social Mind. *Trends in Cognitive Sciences*, *22*(3), 201–212. https://doi.org/10.1016/j.tics.2017.12.005

Thornton, X. M. A., Weaverdyck, M. E., & Tamir, X. I. (2019). The social brain automatically predicts others' future mental states. *Journal of Neuroscience*, *39*(1), 140–148. https://doi.org/10.1523/JNEUROSCI.1431-18.2018

Tigue, C. C., Borak, D. J., O'Connor, J. J. M., Schandl, C., & Feinberg, D. R. (2012). Voice pitch influences voting behavior. *Evolution and Human Behavior*, *33*(3), 210–216.

https://doi.org/10.1016/j.evolhumbehav.2011.09.004

Todorov, A., Pakrashi, M., & Oosterhof, N. N. (2009). Evaluating faces on trustworthiness after minimal time exposure. *Social Cognition*, *27*(6), 813–833. https://doi.org/10.1521/soco.2009.27.6.813

Torre, I., Goslin, J., White, L., & Zanatto, D. (2018). Trust in artificial voices. In *Proceedings of the Technology, Mind, and Society on ZZZ - TechMindSociety '18* (Vol. 18, pp. 1–6). https://doi.org/10.1145/3183654.3183691

Tourville, J. A., & Guenther, F. H. (2011). The DIVA model: A neural theory of speech acquisition and production. *Language and Cognitive Processes*, *26*(7), 952–981. https://doi.org/10.1080/01690960903498424

Van Overwalle, F. (2009). Social cognition and the brain: A meta-analysis. *Human Brain Mapping*, *30*(3), 829–858. https://doi.org/10.1002/hbm.20547

Viscovich, N., Borod, J., Pihan, H., Peery, S., Brickman, A. M., Tabert, M., … Spielman, J. (2003). Acoustical analysis of posed prosodic expressions: Effects of emotion and sex. *Perceptual and Motor Skills*, *96*(3), 759–771. https://doi.org/10.2466/pms.2003.96.3.759

Wang, Y., & Hamilton, A. F. de C. (2012). Social top-down response modulation (STORM): a model of the control of mimicry in social interaction. *Frontiers in Human Neuroscience*, *6*, 153. https://doi.org/10.3389/fnhum.2012.00153

Welch, G. F., Howard, D. M., & Nix, J. (2019). *The Oxford Handbook of Singing*. OUP Oxford.

Weusthoff, S., Baucom, B. R., & Hahlweg, K. (2013). Fundamental frequency during couple conflict: An analysis of physiological, behavioral, and sex-linked information encoded in vocal expression. *Journal of Family Psychology*, *27*(2), 212–220. https://doi.org/10.1037/a0031887

Wild, B., Rodden, F. A., Grodd, W., & Ruch, W. (2003). Neural correlates of laughter and humour. *Brain*, *126*(10), 2121–2138. https://doi.org/10.1093/brain/awg226

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. https://doi.org/10.1111/j.1467-9280.2006.01750.x

Wilson, D. S., Near, D., & Miller, R. R. (1996). Machiavellianism: A synthesis of the evolutionary and psychological literatures. *Psychological Bulletin*, *119*(2), 285–299. https://doi.org/10.1037/0033-2909.119.2.285

Xu, Y., Kelly, A., & Smillie, C. (2013). Emotional expressions as communicative signals. In D. Hancil, S and Hirst (Ed.), *Prosody and Iconicity* (pp. 33–60). John Benjamins Publishing Co. https://doi.org/10.1075/ill.13.02xu

Zuckerman, M., & Driver, R. E. (1989). What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, *13*(2), 67–82. https://doi.org/10.1007/BF00990791

Acknowledgements

Looking back at the last years, I am thankful to many wonderful people who have supported, surrounded and accompanied me, professionally, personally or both.

Carolyn McGettigan – without whom this thesis would not exist – thank you for being an inspiration and an incredible role model, for giving me the chance to work with you at RHUL and being always there despite any circumstance. Thank you!

Herta Flor – who provided me with the opportunity to pursue my PhD under her supervision at the CIMH. You gave me the freedom and confidence to tackle the questions I was passionate about and supported me with your insight and experience.

Frauke Nees – I thank you for your guidance and encouragement. You truly are a superhero, whose door was always open. Thank you for continued help, thoughtfulness and advice.

Thank you to all my colleagues at CIMH and in particular – Kornelius Kammler-Sücker, for being an inspirational colleague and friend. Stefano Silvoni, for always at least one coding solution to any problem and bestowing upon me the delight of Italian baking. Maren Prignitz, for being a fantastic team worker and friend. Sabine Meidner, for your helping hand in any issue big or small, the ever cheerful host of many a happy lunch hour.

I'm also grateful to Elise Kanber and Sheena Walters, who were so welcoming and supportive during my stay at RHUL. And in particular, Nadine Lavan, who never ceases to impress me both professionally and personally.

Thank you to César Lima – whose work sparked my interest into voices, sound and music!

To my fellow PhD students at GESS – especially Sophie Scharf, Mirka Henninger and Liliane Wulff. You have been the most wonderful friends and co-workers from day one. You make this place a home. Thank you also to Milanka Stojkovic, for being incredibly supportive and bringing London to Mannheim with you!

Thank you to Tobias Tröndle, Anne Niemietz and Matthias Sator for reminding me that we are all performers at heart and our voices need to be heard.

And of course, thank you to my family, for helping me get to where I am today each in their own way. I owe so much to you. My mum for her endless belief in me. My uncle, Rainer Süßdorf, for being there for me all the way.

Everything starts and ends with you, Sebastian Siehl, thank you for sharing this life with me. You are my rock, my joy, my home.