UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

THE EFFECT OF SPEAKER'S AGE ON SOCIAL TRAIT INFERENCE IN VOICE PERCEPTION

Marta Almeida e Silva Granadeiro

MESTRADO INTEGRADO EM PSICOLOGIA

Área de Especialização em Cognição Social Aplicada

2019

UNIVERSIDADE DE LISBOA

FACULDADE DE PSICOLOGIA

THE EFFECT OF SPEAKER'S AGE ON SOCIAL TRAIT INFERENCE IN VOICE PERCEPTION

Marta Almeida e Silva Granadeiro

Dissertação orientada pela Professora Doutora Ana Pinheiro e coorientada pelo Professor Doutor Leonel Garcia-Marques

MESTRADO INTEGRADO EM PSICOLOGIA

Área de Especialização em Cognição Social Aplicada

2019

“Sapiens are subject to the same physical forces, chemical reactions and natural-selection processes that govern all living things.”

Yuval Harari

Agradecimentos

À Professora Doutora Ana Pinheiro, por desde muito cedo incentivar a minha investigação e interesse na área do processamento de voz.

À Maria Amorim, pela ajuda incansável e disponibilidade que sempre demonstrou para comigo ao longo deste projeto.

A todos os que, sem dever ou obrigação, contribuíram voluntariamente como participantes para este projeto. Sem isso, nada disto seria possível.

Um especial agradecimento aos meus pais, que serão sempre segundos autores dos meus produtos, por me terem moldado para ser o melhor de mim. Pela educação humana, pelo exemplo e pelo apoio, que para sempre superarão as suas obrigações como pais. Por me serem quem são.

Abstract

When hearing a voice, listeners automatically extract information about the age and infer social traits of the speaker. This information, being accurate or not, defines how we interact with other persons. This study focuses on how speakers from different age groups are perceived. Specifically, how younger adults estimate the age of speakers and how they infer social traits from those speakers. 28 college students estimated the age of child, adolescent, younger adult, adult and older adult speakers. 29 college students rated those voices on dominance and 30 college students rated those voices on trustworthiness, using a 7-point Likert scale. Contrary to previous studies, we did not find an own age bias. Instead we found that children voices are easier to estimate, suggesting a potential role of distinctiveness. Also, not consistent with other studies, our participants had difficulty in accurately estimating the age of speakers. A speaker sex and speaker age group interaction for accuracy in estimating the age from a voice was found, with higher accuracy for male younger adult, adult and older adult speakers. A speaker age and speaker age group interaction was also found for ratings of dominance, reflecting higher dominance ratings for male adolescent, younger adult, adult and older adult speakers. Trustworthiness ratings did not vary as a function of speaker age group. We extend previous findings, from the literature on face perception, to voice perception: the effect of approach/inhibition-related emotions on dominance ratings and the effect of valence of the emotion on trustworthiness ratings. These findings provide useful knowledge with applications in engineering and artificial intelligent systems.

Keywords: voice; age estimation; social traits; perception; social cognition.

Index

Social interaction is something constantly present in our daily lives. We, as a species, manage to live in groups of thousands of individuals, in one place, in cooperation. From the moment you look at the cashier in your local groceries shop, you have already extracted a lot information about him/her. You will have estimated his/her age, sex, how tall he/she is, but also you will infer if he/she is friendly, funny, competent or shallow. We can form an impression of another person in a few seconds, without even talking to them (Willis & Todorov, 2006). As these first impressions are very important in guiding our behaviour, it has become a very important topic of research to understand how we elaborate those impressions and how they shape the way we communicate and interact with each other. Also, we live in a changing world. The world's population reached 7.7 billion people in the middle of 2019 and in 2018, for the first time in our history, people above 65 years old outnumbered those with less than 5 years old, worldwide (Nations, of Economic, Affairs, & Division, 2019). World's life expectancy at birth (meaning the mean number of years a newborn is expected to live, if the death rates do not change) is currently 72.6, which is 8 years more than 30 years ago (Nations et al., 2019). With new age groups emerging, past research is insufficient to completely understand the mechanisms behind age perception and the impact of age on impression formation. The current work is focused on one specific source of social information: the voice. A main goal of the current study was to understand how we infer social traits from the emotional voice of speakers of different ages, as well as identity characteristics, such as age.

Estimating Speaker's Age

Age perception is an important cue for social grouping and consequently it modulates how we respond to and interact with other persons (Hollien, 1987). Voice and face processing share many similarities (Belin, 2017). They are each related to

cortical regions that specifically respond to voice and face stimuli, respectively, and their processing is associated with specific ERP (event-related potential) components. Additionally, a selective deficit in recognition has been found for faces, known as prosopagnosia, (Rossion, 2014) and for voices, known as phonagnosia (Van Lancker, Cummings, Kreiman, & Dobkin, 1988). Also, more recently, a 2D- dimensional space for the perception of social traits, which had been found for faces (Todorov, Said, Engell, & Oosterhof, 2008), was also found for voices (McAleer, Todorov, & Belin, 2014). As there are fewer studies of the perception of age in the voice literature and since there are similarities between the processing of both stimuli (Belin, 2017), models belonging to the face literature will be referred and indirectly used as support for some hypothesis.

There seems to be a replicated effect which is the underestimation of older voices, i.e. older speakers are perceived as of a younger age group, and overestimation of younger voices, i.e. younger adults are perceived as of an older age group ((Shipp & Hollien, 1969; Huntley, Hollien, & Shipp, 1987). Besides this speaker age effect, there is also a listener age effect, as the accuracy in the estimation of a speaker's age varies with listener's age (Huntley et al., 1987; Neiman & Applegate, 1990). One explanation is that listeners seem to be better at recognising the speaker's age when the speaker is of the same age, an effect referred to as own-age bias (Huntley et al., 1987). Consistent with this, younger adult's voices were found to be better recognized in studies testing 20-year-old listeners (Goy, Kathleen Pichora-Fuller, & van Lieshout, 2016; Huntley et al., 1987; Neiman & Applegate, 1990). This effect is also observed when faces are used as stimuli, with children being more accurate at recognizing the age of other children compared to younger adults and adults, older adults being more accurate at recognizing the age of older adults (Anastasi & Rhodes, 2005) and younger adults being more

accurate at recognizing the age of younger adults (Hills & Lewis, 2011). Studies relying on the event-related potential (ERP) technique also support an own-age bias in face perception. Specifically, an increased amplitude of the N170 component and reduced P2 amplitude was observed in children when watching faces of other children, compared to adult and older adult faces (Melinder, Gredebäck, Westerlund, & Nelson, 2010). Increased N170 amplitude in response to faces has been related to an own-race bias, which is expressed as higher accuracy in the recognition of faces of one's own race, and similar P2 amplitudes for own-race faces and other-race faces has been related to *expertise* with other-race faces (Stahl, Wiese, & Schweinberger, 2008). Because of the similar component amplitudes found for adults when watching faces of their own-race in Stahl et al. (2008) and children when watching faces of other children, the latter were interpreted as resulting from *expertise* (acquired with more experience with this age group) in processing the face of people that match their age (in this case, other children) (Melinder et al., 2010). These results support an own-age effect at the electrophysiological level. Hills and Lewis (2011) also found that 7-year-old to 9-year-old children already have an own-age bias.

Motivation and familiarity have been pointed out as possibly driving this effect. The motivation hypothesis states that the higher accuracy in the recognition of the age of speakers who are of the same age of listeners, is due to a higher interest in engaging and interacting with speakers of the same age group (Hills & Lewis, 2011). Familiarity could also play a role as individuals might be more used to respond to their peers, thus, developing *expertise* in recognizing them (Hollien, 1987; Huntley et al., 1987). Also, this could mean that as we get older, the more different age-groups we have dealt with, the more proficient we become at recognising voices from speakers of different ages. This is consistent with studies using faces, showing that older adults are better than

younger adults at discriminating the age of a certain face, while younger adults tend to categorise the faces only as younger or older than themselves (van Rijsbergen, Jaworska, Rousselet, & Schyns, 2014). Considering previous studies, it seems that voices from older speakers are harder to estimate than voices from younger speakers, especially by younger listeners (Shipp & Hollien, 1969; Huntley et al., 1987; Neiman & Applegate, 1990). Despite this rather consistent finding, in the majority of studies, the motivational and the familiarity effects are mixed, since the listener's sample frequently matches the youngest of the speaker's group sample (Goy et al., 2016; Neiman & Applegate, 1990). As we do not know younger adults' performance in estimating the age of younger speakers (e.g. children and adolescents), we cannot tell if the higher accuracy in estimation the age of younger adult speakers is due to younger adult listeners being more motivated in engaging with speakers of their own-age or simply the result of younger listeners being more familiar with speakers with whom they have previously dealt with. Future studies should also include listeners from more age groups.

In the face literature, it seems that in memory tasks, distinctiveness also influences the relationship between the age of a face and the estimator's age. Distinctiveness, which is the relative difference from the average face, i.e. a face prototype derived from the superimposition of several faces, emphasising what is common in those faces and smoothening what is idiosyncratic (van Rijsbergen et al., 2014), has been found to influence face recognition and recall. In fact, Deffenbacher, Vetter, Johanson, & Otoole, (1998) found that ratings of distinctiveness and age had a positive linear relation with distance from an average face, i.e. increasing the distance from the average face, by caricaturing the face, significantly increased perceived age and distinctiveness. Older and more distinct faces generated more correct answers and

were, therefore, considered more memorable. Whether distinctiveness influences the estimation of the age of a face, is not clear. Supporting the influence of distinctiveness, is a study with undergraduate students, in which they were all more accurate at estimating the age of older adult's faces, than younger adult's faces (Bruyer, Mejias, & Doublet, 2007). Although this finding only challenges previous proposals of an own-age bias in faces, considering the similarities between age estimation in faces and in voices (Belin, 2017), it would be important to address the role of distinctiveness in the estimation of age a speaker as well.

Social Trait Inference

Alongside the perception of identity information, such as age, we may also infer social traits from the voice. The perception of social traits from voices has implications on voting behaviour (Tigue, Borak, O'Connor, Schandl, & Feinberg, 2012), leadership choice (Klofstad, Anderson, & Peters, 2012), prosodic entrainment (the changes in speaker's pitch due to interpersonal interaction) (Michalsky & Schoormann, 2017) and financial trust (Montano, Tigue, Isenstein, Barclay, & Feinberg, 2017). In spite of its importance in our communication and behaviour, it is still not clear, exactly how these rapid social trait inferences are made (McAleer et al., 2014). For faces, two dimensions were found, from which the evaluation of social traits is systematically derived: trustworthiness (valence) and dominance (power) (Todorov et al., 2008). Trustworthiness is commonly related to the evaluation of someone as "friend or foe", representing the impression about someone's intentions. Dominance, on the other hand, is commonly related to how capable one is of acting on their intent. Studies using a new technique, called reverse correlation, were able to specifically identify which areas of the face are related to the perception of these two social dimensions (Dotsch & Todorov, 2012). The trustworthy face has been described as being brighter, i.e. more

illuminated, in areas such as the mouth and eyebrow regions and darker, i.e. less illuminated, in areas in the eye and hair regions, whereas dominance was more strongly related to areas around the eyes, hair regions and delineation of the face.

Recently and similarly to the two-dimensional social space for faces, McAleer et al. (2014) proposed a two-dimensional space, for short socially relevant vocal signals, that explained first impression judgements of traits from voices. A Principal Component Analysis (PCA) identified two principal components, corresponding to the traits of trustworthiness/likeability/warmth and dominance/competence. This indicates an identical two-dimensional space for faces and voices, with one dimension related to one's intentions and another with one's ability to act on the intent. Pitch, which represents the mean fundamental frequency of a sound (also known as f0), was the acoustical measure that more accurately predicted social trait representations. However, exactly how pitch is related to the perception of social traits is still unclear. Some studies reported a tendency to evaluate low-pitched voices as more dominant and high-pitched voices as more trustworthy (Ponsot, Burred, Belin, & Aucouturier, 2018), whereas others showed that low-pitched male voices are more dominant and trustworthy, but low-pitched female voices are only more trustworthy (Tsantani, Belin, Paterson, & McAleer, 2016). Also, verbal content seems to influence the role of pitch on social trait perception. In a study of O'Connor & Barclay (2018), low-pitched male voices were assessed as more trustworthy and attractive when the voice had prosocial content (saying words like caring, fair, honest and helpful), but no differences were found for pitch on those traits when the voice had antisocial content (saying words like cheater, fraud, liar and corrupt). The authors propose that the antisocial content of the voice only affected judgements for lower-pitched voices and that the high pitch of male

voices smoothened the antisocial content, thus, not decreasing judgements of trustworthiness and attractiveness for those voices.

Akin to the recent research on face perception, Ponsot et al. (2018) applied reverse correlation to voice samples in order to better understand which specific prosody of speech relates to the perception of a given social trait. In this study, participants listened to two voices saying “bonjour” (“hello” in English) and had to choose which one was more dominant/trustworthy. A negative correlation between dominance and mean pitch and a positive correlation between trustworthiness and mean pitch was replicated, for both male and female listeners. Using reverse correlation they were able to construct a voice prototype for each main dimension (dominance and trustworthiness), these prototypes resulted from summing the pitch patterns that were rated as more dominant/trustworthy and subtracting the pitch patterns that were rated as less dominant/trustworthy. The prototypes are thought to reflect the mental representation of a dominant and trustworthy voice. After that, by analysing the dynamic pitch contours of the dominant voice prototype, they found a linear relationship between pitch and perceived dominance, with the dominant prototype having a gradual decrease in pitch over the two syllables of the word “bonjour”. However, trustworthiness and pitch had a more complex relation. Perceived trustworthiness was predicted by the pitch gap between the first and second syllable of the French word “bonjour”, specifically the rapid increase of pitch on that gap. Additionally, applying the trustworthiness dynamic pitch pattern (the trustworthiness prototype) to other two-syllable words did not increase perceived trustworthiness for those voices, showing that the perception of social traits cannot be explained by simple general pitch variations (Ponsot et al., 2018). Other acoustic measures of the voice have proven to be important for the perception of social traits, for instance, harmonic-to-noise

ratio (HNR), indicating roughness, intonation (changing f0) and glide (difference between f0-end and f0-start) (McAleer et al., 2014). So, although pitch is probably the most studied acoustical measure, the perception of social traits cannot be explained only by pitch, as it is a result of the integration of many acoustic signals (McAleer et al., 2014).

The decoding of age information and social traits from the voice plays a critical role in social communication, impression formation and behaviour (Huntley et al., 1987; Klofstad et al., 2012; Michalsky & Schoormann, 2017; Montano et al., 2017; Tigue et al., 2012). For faces, the resemblance of a visual cue (e.g. having larger eyes and rounder face) with those corresponding to the mental representation of an age or trait (e.g. babies/low power), will make it more probable for the face to be perceived as having that age or trait (e.g. he/she is young/has low power) (Zebrowitz, 2017). This overgeneralisation observed in social trait inference and age estimation for faces, is also congruent with finding for voices, as for instance, low-pitched voices are usually perceived as more dominant (Ponsot et al., 2018). The perception of a low pitch in a voice (acoustical cue), resembles that of a dominant person (trait mental representation), therefore, will result in inferring that the person is dominant (overgeneralisation). Another example would be, perception of a high-pitched male voice (acoustical signal), resembles that of an older adult male speaker (mental representation) (Hollien, 1987) and will more probably be identified as an old man.

For faces, the perception of dominance and age appear to have an inverted-U shape relation. Dominance ratings increase till approximately 35 years old and decrease after that, with increasing age (Batres, Re, & Perrett, 2015). However, because age was significantly correlated with masculinity but not with dominance alone, these authors, suggest that age is correlated with masculinity (the difference from the average female

prototype), which in turn is correlated with dominance. That is, the age-dominance relationship is mediated by masculinity (Batres et al., 2015). However, if the relationship between dominance and age has an inverted-U shape, a linear correlation (used in this study) would not reflect the true nature of the relation and would result in a very low correlation between age and dominance. Accordingly, a quadratic model (compared to a linear model) was found to better fit the relation between age and dominance (Batres et al., 2015).

Regarding trustworthiness, for voices, there seems to be a speaker age effect, as younger voices are rated as more trustworthy than older voices (Schirmer, Feng, Sen, & Penney, 2019). There are also suggestions of a listener's age effect, that is, older listeners' ratings of trustworthiness of a voice were marginally higher than those of younger listeners (Schirmer et al., 2019). Besides the speaker and listener age effects, there also seems to be a speaker's sex effect for trustworthiness, with women being rated as more trustworthy than men (O'Connor & Barclay, 2018; Schirmer et al., 2019). Although some conclusions have been drawn, a model capable of unifying these results, does not yet exist. Furthermore, age does not seem to have a clear influence on the perception of some social traits, such as dominance and trustworthiness, whereas it has a linear impact on others, namely attractiveness, that seems to decrease with increasing age (Deffenbacher et al., 1998), regardless of the rater's age group (young or old) (belonging to the face literature, see Kiiski, Cullen, Clavin, & Newell, 2016).

Interface Between the Emotional Voice, Speaker Age and Social Trait Inference

Another important aspect of voice processing is the emotionality of the voice, which we have to rapidly assess in our daily social interactions. The existing literature suggests that aging is associated with some physical changes in the vocal tract, as well as in the hearing system (Hollien, 1987). Those changes may affect voice perception

and production over the years. First, acoustic analysis and subjective reports show that older voices are usually perceived as slower (Shipp, Qi, Huntley, & Hollien, 1992) louder and less pleasant (Goy et al., 2016). These characteristics may make it more difficult to decode the emotion and age of older voices. Second, the deterioration of the hearing apparatus over the years may also make it more difficult for older listeners to process auditory stimuli, especially high frequency ones (Hollien, 1987). Hence, it is not surprising that younger listeners outperform older listeners when assessing the emotional category of a vocal stimuli (Sen, Isaacowitz, & Schirmer, 2018). This difference (younger listeners better performance compared to older listeners) is smaller when the voice is of an older speaker compared to a younger speaker, which is also not surprising if we consider that older speaker's emotional expression is altered, making it harder to decode its emotionality. So, when it comes to younger speakers, older listeners' hearing difficulties may be responsible for the difficulty in perceiving emotion, nevertheless, when it comes to older speakers, both younger and older listeners have difficulty in perceiving emotion.

Importantly, besides de listener's age effect, it seems that emotion category also influences this effect. For older listeners, the recognition of emotions of younger speakers was comparable to that of older speakers when the emotion assessed was happiness, but not when it was anger, sadness and neutral expression (Sen et al., 2018). This is in line with the selectivity theory, which postulates that with the perception of time remaining in life comes a preference for short-term goals and rewards (Carstensen, 1995). This might emerge as a preference or attentional bias to more positive information (Bailey, Slessor, Rieger & Rendell, 2015). In a study comparing age-related differences in a trust game, older adults were more prone to invest in trustees that were uncooperative and provided low returns than younger adults (Bailey et al., 2015). This

finding is also in line with the dynamic integration theory (Labouvie-Vief, 2003), which postulates that positive stimuli consume less resources than negative ones. Regarding dynamic integration theory however, neutral stimuli, which have less information to process (no emotional charge), should also be easier for older listeners to process. Nevertheless, in the study of Sen et al. (2018) older listeners were significantly worse than younger listener in recognising neutral stimuli, a result similar to that of negative emotions (whose processing is supposedly more cognitively demanding). This challenges this theory's explanation of the age differences in emotion recognition.

The relationship between the emotional quality of the voice and the perception of social traits remains to be specified. Will the emotion expressed by a speaker change how the listener infers social traits of the speaker? When assessing trustworthiness, the perceived positive valence of the stimuli positively predicted perceived trustworthiness (Schirmer et al., 2019). This is thought to support the emotion metaphorical generalisation, which postulates that when social information that is important for forming an accurate impression of someone is missing (for example, previous behaviour or pattern of relationship), we tend to fill that gap in a way that is congruent with our cognitive structure (Secord & Stritch, 1960). Hence, if when we listen to a speaker, we do not hold sufficient information to assess that person, we will organise the information we have in a way that allows us, for example, to predict his or her behaviour with some degree of certainty. So, the expression of a positive emotion would signal a probable positive behaviour, which in turn would increase the expectation of a good intention, eliciting more trust (Schirmer et al., 2019; for face literature also Zebrowitz, 2017; Zebrowitz & Montepare, 2008). Negative emotion, on the contrary, would signal a probable negative behaviour, which would increase the expectation of a bad intention and elicit less trust. The generalisation of emotional valence to a more

stable trait (trustworthiness) has also been found in children as young as 10 years old (Caulfield, Ewing, Bank, & Rhodes, 2016).

The metaphorical generalisation (Secord & Stritch, 1960) also holds true for dominance, as the expression of some emotions is related to higher ratings of dominance (Keltner, Gruenfeld, & Anderson, 2003), suggesting that they signal a behaviour congruent with that social trait. Whilst trustworthiness seems to have a linear relationship with emotional valence, dominance does not. Angry faces were associated with higher dominance ratings, followed by happy faces, followed by neutral faces and, finally, by sadness (Sutton, Herbert, & Clark, 2019). However, Hareli, Shomrat, & Hess (2009) found no difference in ratings of dominance between angry and neutral faces in male faces. In the case of dominance, it has been hypothesised that emotions related to the approach system (such as anger, enjoyment and pleasure) are related to higher dominance ratings, whereas emotions related to the inhibition system (such as fear and sadness) are related to lower dominance ratings (Keltner et al., 2003). This approach system is the result of behavioural and social consequences of having higher power, for example, it has been observed that individuals with more power show more attention to reward, more disinhibited behaviour, more positive emotion expression and experience, and more simplistic and automatic appraisal of the social environment. The inhibition system, on the other hand, which holds the behavioural and social consequences of having lower power, includes more attention towards punishment and threat, experiencing and expressing negative feelings and more inhibited behaviour (Keltner et al., 2003). It would be expected of someone with more power to express more anger when frustrated or when failing, to express more happiness when satisfied and to express more pleasure in a pleasant situation, whereas someone with lower power would for example express guilt when failing and less pleasure in a pleasant situation

(Hareli et al., 2009; Keltner et al., 2003). As dominance and power are two very closely related (and often overlapping) social concepts (Todorov et al., 2008), it would be expected that the approach and inhibition-related emotions would behave in a similar fashion for both traits. Hareli et al. (2009) proposed that the approach-related emotions (such as anger and happiness, for instance) do not increase perceived dominance, but instead the inhibition-related emotions (such as fear or sadness) decrease it. However, they did not replicate this finding for women's faces: neutral and fearful female faces were rated as lower in dominance, compared to angry and happy faces. This sex difference may have not been accounted by in the study of Sutton et al. (2019), which may have led to discrepant ratings of dominance for emotional vs. neutral stimuli. As there seems to be discrepant results on the role of approach vs inhibition system on social trait perception, including neutral stimuli is especially important. First to have a control condition in order to understand the real contribution of the different emotions, second, because in some contexts neutral stimuli might actually be a sign of dominance, in a way that it shows confidence and self-assurance when dealing with a stressful event (Hareli et al., 2009).

The Current Study

The current study is aimed at further understanding how younger adults estimate the age of different age groups, as well as how age relates to the perception of certain social traits. Firstly, and as mentioned before, speaker's age and listener's age both influence the perceived age of a voice, the result of this interaction has been called the own age bias. This own age bias might be the result of either familiarity or motivation. To disentangle these possibilities, we included children and adolescent, as well as younger adult, adult and older adult speakers. By including these age groups, not only can we study how younger adult listeners perceive the age of those younger than

themselves, but also directly compare with how they perceive voices of those older and of the same age as themselves. If it is motivation leading to higher accuracy in the estimation of age, younger adult listeners should be better at estimating the age of younger adult speakers (who match their age) than of any other age group. If instead, it is familiarity leading to higher accuracy in age estimation, younger adult listeners should be better at estimating the age of younger adults, as well as of children and teenagers, because they have previous experience with those age groups, hence, are more familiar with them.

The Belin, Fecteau, and Bédard (2004) model for voice processing suggests that there is functional independence between the processing of identity information and emotional information of the voice. However, these processes are not fully independent and brain areas related to each one of those processes are likely to interact. For this reason, in this study we assessed if emotion plays a role (facilitating or making it more difficult) in the perception of identity information from the voice, in this case, the estimation of age. However, we do not predict any specific magnitude or nature of this possible effect.

Secondly, because the same perceptual cues signal certain age groups and certain social traits, we expect an overgeneralisation, meaning if the voice's perceptual cues resemble those of a specific age or trait, that voice will be perceived as having that age and trait. Specifically, as dominance is positively related to access to resources (e.g. material resources such as money, or political resources such as decision making), physical strength, power (Keltner et. al, 2003), height, masculinity (Batres et al., 2015) and competence (Todorov et al., 2008), voices sharing vocal cues with the above-mentioned prototypes will be perceived as more dominant. On one hand, for male faces, height ratings had a positive relation with masculinity and an inverted-U shape relation

with age (i.e. height ratings increased till middle age but decreased after that) (Batres et al., 2015). For voices, male children and older adults' voices are known to be less masculine (Hollien, 1987). Therefore, it is expected that male children and male older adult voices will be perceived as less dominant. On the other hand, male adolescents, male younger adults and male adult's voices are known to be more masculine (Hollien, 1987), as well as being perceived as taller than male children and male older adults (this latter result is from the face literature) (Batres et al., 2015). So, it is expected that male adolescents, male younger adults and male adult speakers, will be perceived as more dominant. Importantly, because female voices seem to have less acoustical changes over the years (compared to male voices) (Hollien, 1987), a speaker age and speaker sex interaction is also expected, with adolescent, younger adult and adult male speakers being perceived as significantly more dominant than adolescent, younger adult and adult female speakers.

Concerning trustworthiness, older adult speakers are perceived to be less trustworthy than younger adult speakers and male speakers as less trustworthy than female speakers (Schirmer et al., 2019). Additionally, for faces, having perceptual cues that resemble those of a baby, elicits childish-like traits, such as low competence, low power, high warmth (Zebrowitz & Montepare, 2008) and high trustworthiness (Li, Heyman, Mei, & Lee, 2019). Considering this, we expect older speakers to be perceived as less trustworthy and children, adolescent and younger adult speakers as more trustworthy, as well as female speakers to be perceived as more trustworthy than male speakers.

Thirdly, the perception of different emotions, like that of different social traits, is related to the processing of perceptual cues that resemble those belonging to a certain emotion, i.e. if a voice has acoustical parameters that resemble those of a voice

expressing happiness, it will be perceived as expressing happiness. Moreover, not only will that voice be perceived as conveying happiness, but also as probably being from someone who is trustworthy and a female (Schirmer et al., 2019; see also Adams, Nelson, Soto, Hess, & Kleck, 2012 for support from the face literature). This metaphorical generalisation (Secord & Stritch, 1960) is thought be responsible for the relation between the perception of emotion and inference of social traits, and to happen irrespective of age (Zebrowitz & Montepare, 2008, also from the face literature). Furthermore, the expression of specific emotions has been associated with the perception of specific social traits (Hareli et al., 2009; Keltner et al., 2003; Schirmer et al., 2019; Sutton et al., 2019). The approach and inhibition-related emotions have been related to the perception of dominance (Hareli et al., 2009; Keltner et al., 2003, from the face literature), and valence of the emotion has been related to the perception of trustworthiness (Schirmer et al., 2019; see also Sutton et al., 2019; Caulfield et al., 2015 from the face literature).

In this study we included voices expressing anger, enjoyment, pleasure, disgust and neutral, for two reasons. First, because these emotions had the highest recognition scores in a pilot study (accuracy ratings above 55%) and we wanted to make sure the manipulation of emotion category was effective. Second, two emotions have negative valence and two emotions have positive valence, which allows us to explore the overgeneralisation occurring for trustworthiness judgments, also we included neutral as a control (Hareli et al., 2009). Regarding trustworthiness, we expect that voices expressing enjoyment and pleasure will be perceived as more trustworthy than voices expressing anger and disgust. More specifically, if it is the positive valence of the voice driving higher trustworthiness ratings, then it is expected that trustworthiness of neutral voices will be different from happy and pleasure but not from anger and disgust. If it is

the negative valence of the voice driving higher trustworthiness ratings, trustworthiness ratings for neutral voices will be different from angry and disgust, but not enjoyed and pleasured. If both, negative and positive valence, relate to trustworthiness ratings, then happy and pleasure voices will be perceived as the most trustworthy and neutral as more trustworthy than anger and disgust. Regarding dominance, if the approach/inhibition-related system is what drives judgements of dominance, it would be expected that voices expressing anger, enjoyment and pleasure would be rated as more dominant than voices expressing disgust (Hareli et al. 2009; Keltner et al., 2003; Sutton et al., 2019). Specifically, if it is the presence of inhibition-related emotions that leads to lower perceived dominance, we would expect neutral voices to be perceived as more dominant than those expressing disgust, and not differing from the ones expressing anger, enjoyment and pleasure (Hareli et al., 2009). If it is the presence of approach-related emotions that increases perceived dominance, we would expect neutral voices to be perceived as less dominant than those expressing anger, enjoyment and pleasure, but not different from disgust. If both systems are responsible for the perceived dominance, we would expect anger, enjoyment and pleasure to be rated as more dominant than neutral and disgust, and neutral as more dominant than disgust (Sutton et al., 2019). We do not expect any interaction with age nor sex (Zebrowitz, 2017).

Method

Stimuli

Vocalisations produced by 20 speakers from five different age groups (8-11 years; 14-16 years; 19-23 years; 40-50 years; >60 years), 10 females and 10 males (2 of each sex, per age rank) were recorded. All speakers signed an informed consent allowing the recording of their voices. For the participation of children and adolescents, the written consent was given by their parents. An Edirol R-09 recorder and Shure

PG48 microphone was used to record the voices. Younger adults were recorded in an anechoic chamber at the University of Minho and due to availability constraints, vocalisations produced by children, adolescents, adults and older adults were recorded in a quiet room (for example primary school and senior university).

The vocalisations were elicited by watching a video and reading a scenario aiming to elicit different emotions. Participants were instructed to be spontaneous, natural and brief, and to utter emotional sounds without verbal content that were congruent with the emotional situation they saw/read (Lima, Castro, & Scott, 2013). They were also asked to utter different emotions expressing with the vowel "a" (Lima et al., 2013). For this study, we only selected two positive emotional categories (amusement and pleasure), two negative emotional categories (anger and disgust) and neutral vocalisations. All categories had a recognition accuracy of more than 55%. This resulted in 100 different stimuli.

Background noise was removed using version 2.1.1 of Audacity software (®) (Audacity Team, 2014) intensity was normalised (to 70dB) using Praat software (version 5.1.05; Boersma & Weenik, 2009). Physiological sounds, such as sneezing, hiccup, verbal interjections, were also removed (Schroder, 2003), as well sounds with recording problems. This was meant to ensure that the stimuli had the appropriate acoustic parameters.

Task 1: Estimation of Speaker's Age

Participants. Twenty-eight participants were included in the age task (26 females and 2 males, mean age=20.04 years; SD=1.538 years). All participants reported normal hearing, no neurological impairments and were recruited through social networking sites. Participants provided consent after reading the informed consent and

were told they could ask any question they needed during the experiment. They could not proceed to the experiment without having given their consent. The study was approved by the Ethics Committee of the Faculty of Psychology of the University of Lisbon.

Procedure. Participants completed the experiment in group sessions (maximum 7 people) using headphones. They were told that they would hear vocalisations and, after each one, had to estimate the age. They were also told that they could hear the vocalisation as many times as they needed and to inform, the experimenter, if any volume adjustment was needed. Information on their age, sex, course name and course year was collected. All participants provided informed consent (by pressing [SPACE], otherwise they could not continue the experiment). After this, they read the instructions and started with a four-trial training to get familiarised with the task, and afterwards completed the rest of the task.

Each trial began with a fixation cross for 500ms, followed by a vocalisation. They were asked to estimate the age group of the speaker by pressing one of the numbers **1** (8-12 years), **2** (14-17 years), **3** (19-25 years), **4** (40-50 years) or **5** (+60 years), corresponding to the age group of the speaker. There was a 2000ms interval between the answer and the next stimulus (Figure 1). The age, sex and emotion category of the voice were randomised.

Vocalisations were presented via E-Prime 2.0. (Psychology Software Tools, Pittsburgh, PA) through headphones. Each participant listened to 100 vocalisations. The task took approximately 20 minutes.

Task 2: Estimation of Speaker's Social Traits

Participants. Thirty participants were included in the dominance task (18 females and 12 males, mean age=20.93; SD=1.41; age range = 19-23) and twenty-nine in the trustworthiness task (15 females and 14 males, mean age=21.10 years; SD=2.11 years; age range =18-26). All participants reported normal hearing, no neurological impairments and were both recruited through social networking sites. One participant had a ruptured tympanic membrane in one ear and used one earphone but did not report any difficulty in performing the task. Participants provided formal consent after reading the informed consent and were told they could ask any question they needed during the experiment. They could not proceed to the experiment without having given their consent. The study was approved by the Ethics Committee of the Faculty of Psychology of the University of Lisbon.

Procedure. The social trait task was performed in a similar fashion, except participants were told that after each vocalisation they would have to rate the voice in a 7-point Likert scale, ranging from 1 (not at all [trait]) to 7 (extremely [trait]). They were also told that they could hear the vocalisation as many times as they needed. Participants assigned to the dominance task did not perform the trustworthiness task and vice versa.

The social trait task had the same structure as the age task, but instead of estimating the age of a voice, participants had to rate how dominant/trustworthy the voice was, by pressing a number from **1** to **7** (Figure 1). The task took approximately 20 minutes.

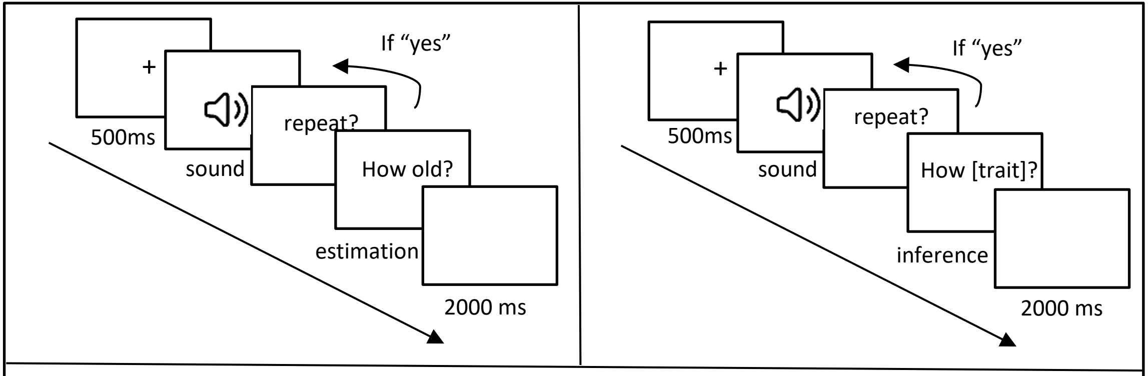


Figure 1. Experimental trial for each task. Age estimation (left) ranged from 1 to 5. Social trait inference (right) ranged from 1 to 7.

Statistical Analysis

All statistical analysis was conducted using IBM SPSS Statistics for Windows, version 25.0 (IBM Corp., Armonk, N.Y., USA).

Firstly, to assess inter-rater agreement, an Intraclass Correlation Coefficient (ICC) was calculated for each task (Schirmer et al., 2019). In the social trait inference task, a One-way ANOVA was conducted in order to asses if there were differences between male and female raters. This was not performed for the age estimation task, as the participants were almost all female. As our listeners were of the same age and revealed no other relevant differences between them, and, because the focus was how the different stimuli were perceived, each stimulus was used as unit of analysis.

For the age estimation task, absolute hits were transformed into hit rates, by dividing absolute hits by the total of answers to each stimulus. So, for instance, a hit rate of 0.50 would mean that half the responses to given stimulus were correct. However, group hit rates are the result of the average of hit rates of each stimulus of that group, i.e. the hit rate for male voices would be the result of averaging hit rates of all stimuli of

male speakers. That is, a hit rate of 0.50 for male voices, would mean that half the estimations of the age of male speakers were correct. If there is high agreement between listeners, using averaged hit rates, instead of individual hit rates, may be advantageous, as it smoothens individual variance, making it a robust measure. To assess if hit rates varied between speaker age group, speaker sex, and emotion category expressed, a multifactorial ANOVA was used. This resulted in a 5 (age group: Children vs. Adolescents vs. Younger Adults vs. Adults vs. Older Adults) x 2 (sex: female vs. male) x 5 (emotion category: neutral vs. anger vs. disgust vs. enjoyment vs. pleasure), between-subject design.

Because we were not only interested in the amount of hits/errors, but also the type of errors made (underestimation errors or overestimation errors), we qualitatively analysed the distribution of hits and errors for each speaker age group. So, for each speaker age group (e.g. adults), we analysed the percentage of correct (e.g. adult) and incorrect (e.g. children, adolescent, younger adults and older adult) responses. This way we could specifically know if certain age groups are mainly being underestimated or overestimated.

For the social trait inference task, ratings for each stimulus were also averaged. To assess if dominance and trustworthiness ratings varied with speaker age group, speaker sex and emotion category, a multifactorial ANOVA was conducted for each trait. The result was a 5 (age group: Children vs. Adolescents vs. Younger Adults vs. Adults vs. Older Adults) x 2 (sex: female vs. male) x 5 (emotion category: neutral vs. anger vs. disgust vs. enjoyment vs. pleasure) between-subject design. Additionally, a curve estimation analysis (see Bartes et al., 2015) was done to compare if the relationship between social trait inference (dominance or trustworthiness) and speaker

age versus social trait inference (dominance or trustworthiness) and perceived speaker age, differed.

Results

Task 1: Estimation of Speaker's Age

An ICC of 0.98 was obtained, which shows that participants highly agreed on their estimations (Appendix A).

A multifactorial ANOVA was conducted with speaker age group, speaker sex and emotion as between-subject variables, resulting in a 5 (age group: Children vs. Adolescents vs. Younger Adults vs. Adults vs. Older Adults) x 2 (sex: female vs. male) x 5 (emotion category: neutral vs. anger vs. disgust vs. enjoyment vs. pleasure) design. There was a main effect of sex, F (1, 50) =11.047, p=0.002, partial η^2 =0.181, showing higher hit rates for male voices compared to female voices. A main effect of age group was also observed, F(4, 50)=6.045, p<0.001, partial η^2 =0.326. Pairwise comparisons with Bonferroni correction revealed that children's voices were significantly better estimated than adolescents'(p=0.001), younger adults' (p=0.026) and older adults' voices (p=0.002). There was no difference between adult's voices and any other age-group and no other differences between age groups (p>0.05).

An interaction with speaker age group qualified the speaker sex main effect, F (4, 50) =2.606, p=0.047, partial η^2 =0.173. Whereas for younger adult (p=0.018), adult (p=0.012) and older adult speakers (p=0.016), the age of male speakers was better estimated than that of female speakers, no differences for sex were found for children (p=0.296) nor adolescent speakers (p=0.248) (Figure 2). The analysis also showed a speaker age and emotion category interaction, F (16, 50) =1.958, p=0.036, partial η^2 =0.395. The age of children speakers expressing pleasure was better estimated than

that of adolescent (p=0.002) and older adult speakers (p=0.001) expressing pleasure. In the case of neutral vocalisations, the age of children speakers was better estimated compared to older adult speakers (p=0.001).

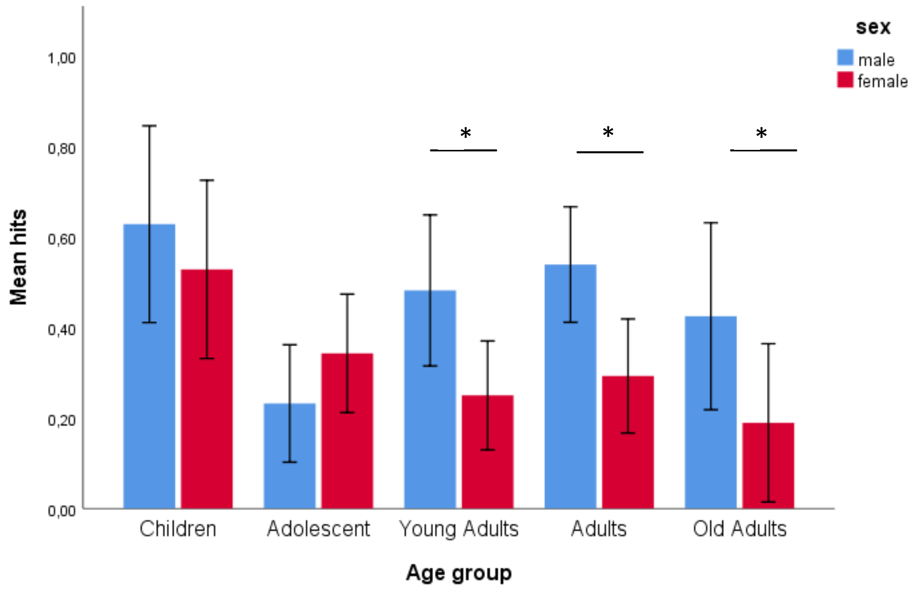


Figure 2. Mean hit rates for each age group and sex. Bars represent 95% Confidence Interval (CI). *p<0.05

The distribution of hits and errors in the estimation of age, for each specific age group was also examined. As shown in Table 1, almost 30% of errors in estimating the age of children speakers resulted in the perception of those voices as belonging to adolescent speakers. For adolescent speakers, the percentage of overestimation (a result of adolescent speakers being perceived as younger adult speakers - 31.6%) was larger than the percentage of hits (28.9%). Regarding younger adult speakers, they were both underestimated (24.1% errors were the result of being perceived as adolescent speakers) and overestimated (23.8% errors were the result of being perceived as adult speakers). Adult speakers were more underestimated (27.5% errors were the result in being

perceived as younger adult speakers) than overestimated (17.7% errors were the result of being perceived as older adult speakers). Older adult speakers had a high percentage of underestimation errors (42%), as a result of being perceived as adult speakers. Except for children speakers, and although the estimation of age was above chance for all age-groups (chance level would be 20% hits), the percentage of errors ranged from 58% to 70%.

Table 1

Distribution of participants' estimation of age for each speaker age group

	Estimated age group									
	1		**2**		**3**		**4**		**5**	
Age group	N	(%)	N	(%)	N	(%)	N	(%)	N	(%)
1	324	(**57.9**)	151	(27.0)	53	(9.5)	18	(3.4)	14	(2.5)
2	76	(13.6)	162	(**28.9**)	177	(31.6)	128	(22.9)	17	(3.0)
3	77	(13.8)	135	(24.1)	205	(**36.6**)	133	(23.8)	10	(1.8)
4	7	(1.3)	68	(12.1)	154	(27.5)	232	(**41.4**)	99	(17.7)
5	11	(2.0)	34	(6.1)	107	(19.1)	235	(42.0)	173	(**30.9**)

Note. N = absolute hits and errors; 1 = children; 2 = adolescent; 3 = younger adult; 4 = adult; 5 = older adult. Percentage of hits for each age group in bold.

Task 2: inference of speaker's social traits

Dominance. Participants showed high agreement in their evaluation of the dominance of the different vocalisations, reflected in an ICC (Intraclass Correlation Coefficient) of 0.886 (see Appendix A). Male and female participants did not differ in

their ratings, Z (1) =0.576, p=0.454. Therefore, the analysis proceeded with male and female ratings together.

A multifactorial ANOVA was conducted. Speaker age group, speaker sex and emotion category were included as between-subject variables resulting in a 5 (age-group: Children vs. Adolescents vs. Younger Adults vs. Adults vs. Older Adults) x 2 (sex: female vs. male) x 5 (emotion category: neutral vs. anger vs. disgust vs. enjoyment vs. pleasure) design. The ANOVA revealed a main effect of age group, F (4, 50) =16.923, p<0.001, partial $\eta^2 = 0.575$. Post-hoc comparisons with Bonferroni correction revealed that vocalisations produced by children were perceived as less dominant than vocalisations produced by adolescents (p=0.001), younger adults (p<0.001) , adults (p<0.001) and older adults (p<0.001); vocalisations produced by adults were rated as significantly more dominant than vocalisations produced by children (p<0.001) and adolescents (p=0.037), but not than vocalisations produced by younger adults (p=0.125) or older adults (p=1). There was also a main effect of sex, with male voices being rated as more dominant than female voices, F (1, 50) = 61.530, p<0.001, partial $\eta^2 = 0.552$. Importantly an interaction of speaker age qualified the speaker sex main effect, F (4, 50) =2.546, p=0.033, partial $\eta^2 = 0.185$ (Figure 3; see also Appendix B), such that male adolescent (p<0.001), male younger adult (p<0.001), male adult (p=0.01) and male older adult speakers (p=0.001) were considered more dominant than female speakers of those age groups. In the case of children speakers, there were no differences in attributed dominance as a function of speaker sex (p=0.239). There was also a main effect of emotion category, F (4, 50) = 7.555, p<0.001, partial $\eta^2 = 0.377$ (Appendix C and D). Post-hoc comparison with Bonferroni correction revealed that voices expressing anger were perceived as significantly more dominant than those expressing disgust (p<0.001), pleasure (p=0.003) and neutral ones (p<0.001), but not enjoyment (p=0.384).

Vocalisations expressing enjoyment did not differ from any other emotional category (*p*>0.05 for anger, disgust, pleasure and neutral) and the comparisons between the other emotional categories were also not significant (*p*>0.05 for all comparisons).

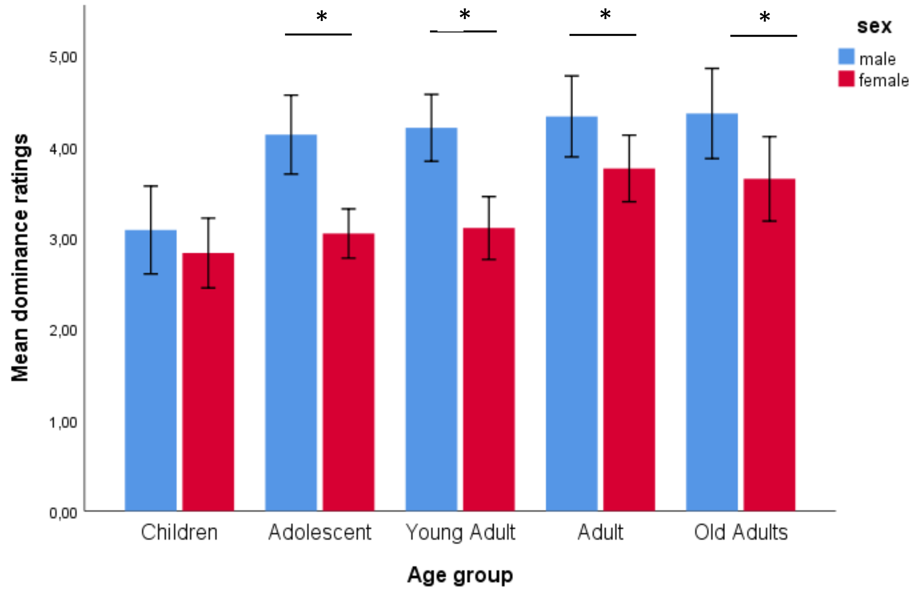


Figure 3. Mean dominance ratings for each age group and sex (*p<0.05).

A curve estimation analysis was performed for the relation between dominance ratings and chronological age, and dominance ratings and perceived age. This revealed that when we consider the real age of the speaker, the quadratic model represented a slightly better fit than the linear model (adjusted R^2=0.22 and adjusted R^2=0.20, respectively; Appendix E). However, when we consider the perceived age of the speaker, the quadratic model does not explain additional variance compared to the linear model (R^2 =0.620 and R^2=0.624, respectively). This suggests that the relation between dominance ratings and the age of the speaker may vary depending on our variables being the real (or chronological) age or the perceived age of the speaker.

Trustworthiness. As in the dominance ratings, there was high agreement between raters, with an ICC of 0.86 (Appendix A). Again, no differences between male and female raters were observed, Z (1) =0.211, p=0.649. A multifactorial ANOVA was conducted with speaker age group, speaker sex and emotion category as between-subject variables, resulting in a 5 (age-group: Children vs. Teenager vs. Younger Adults vs. Adult vs. Older Adults) x 2 (sex: female vs. male) x 5 (emotion category: neutral vs. anger vs. disgust vs. enjoyment vs. pleasure) design. The ANOVA revealed a main effect of sex, F (1, 50) =4.190, p=0.046, partial η^2 =0.077, indicating that female voices were perceived as significantly more trustworthy than male voices (Appendix F). A main effect of emotion was also found, F (4, 50) =24.803, p<0.001, partial η^2 =0.665 (Figure 4; Appendix D). Post-hoc comparisons with Bonferroni correction revealed that vocalisations expressing anger were perceived as less trustworthy when compared to vocalisations expressing pleasure(p<0.001), enjoyment (p=0.001) and neutral voices(p<0.001), but not disgust (p=1). Vocalisations expressing pleasure were perceived as significantly more trustworthy than vocalisations expressing enjoyment (p=0.046) but not neutral voices (p=1) . Vocalisations expressing enjoyment did not differ from neutral ones (p=0.338). No other significant effects were observed (p>0.05 for all comparisons).

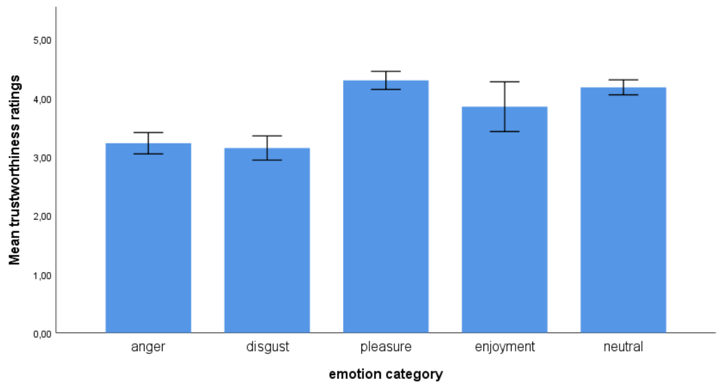


Figure.4 Mean trustworthiness ratings for each emotion category. Bars represent 95% CI.

Trustworthiness ratings did not vary as a function of the real age of the speaker. However, they might be affected by the perceived age, so a curve estimation analysis was performed for trustworthiness ratings and perceived age. Both the linear and quadratic models explained very little variance, and neither was a good fit for the relationship between trustworthiness and perceived age (adjusted R^2=0.059 and adjusted R^2=0.058, respectively).

Discussion

This study investigated the estimation of the age and inference of social traits from vocalisations produced by speakers of different ages and conveying different emotions. It was expected that the age of children, adolescent and younger adult speakers would be better estimated, as a result of an own age bias and familiarity with age groups with whom they have previous experience (Huntley et al., 1987; Neiman & Applegate, 1990; Shipp & Hollien, 1969). Our results did not support an own age bias in the estimation of age, instead it is suggested that the acoustical distinctiveness of children's voices made it easier to estimate their age. Adolescent speakers and older adult speakers were mostly perceived as younger adults and adults, respectively. The age of male speakers was more accurately estimated in all speaker age groups, except children speakers. This is in line with the male voice having more age-related changes in acoustical parameters (e.g. fundamental frequency). The estimation of the age of speakers was not the same for all emotions, this contradicts functional independence between the processing of identity information and emotional information of the voice (Belin, 2004). This extends previous findings and questions the own age bias found in previous studies (Huntley et al., 1987; Neiman & Applegate, 1990; Shipp & Hollien, 1969). The high accuracy in estimating the age of speaker is also questioned, as the

percentage of correct estimations of age was quite low for all speaker age groups, except children.

Furthermore, the overgeneralisation hypthesis (Zebrowitz, 2017; Zebrowitz & Montepare, 2008) predicted the observed results for the perception of trustworthiness and dominance and its relation to age and emotion. Specifically, dominance ratings were higher for male adolescent, younger adult, adult and older adult speakers, compared to female speakers of those age groups, but not for children. Dominance ratings were also higher for speakers expressing anger and enjoyment, as expected by the approach/inhibition-related system (Hareli et al., 2009; Keltner et al., 2003). Trustworthiness ratings did not vary as a function of speaker age, however, trustworthiness ratings were higher for speakers expressing neutral and positive emotions than for speakers expressing negative emotions (Sutton et al., 2019). Our results also suggest that, although this general trend (dominance and trustworthiness ratings relation to emotion) was in accordance with previous findings, a more specific emotion categorisation would probably be more adequate in explaining the relation between emotion and social trait inference. This study suggests similarities, in the inference of social traits, with the literature on face perception (Belin, 2017).

Is It an Own Age Bias?

Firstly, and not consistent with other studies (Huntley et al., 1987; Neiman & Applegate, 1990; Shipp & Hollien, 1969), our participants showed difficulties in estimating the age of the voice. Importantly, the age of children speakers was the easiest to estimate, followed by adult speakers, which does not suggest an own age bias (i.e. listener's higher accuracy in estimating the age of speakers that match their own age). Older adults' voices were as hard to estimate as adolescents', younger adults' or adults',

which is also not in line with previous reports of lower accuracy in estimating the age of older speakers, especially for younger adult listeners (Huntley et al., 1987; Neiman & Applegate, 1990; Shipp & Hollien, 1969). Motivation could be an explanation for the higher accuracy in estimating the age of children speakers, as their voices would trigger the need for assistance or care giving (Little, 2012). However, it would be expected of younger adults to be more motivated in processing younger adult's voices as well (Hills & Lewis, 2011), which was not observed here. Familiarity also does not seem to explain why younger adults were better at estimating the age of children. It is unlikely that younger adults are more exposed to or used to dealing with children's voices, compared to other age group (they may have younger siblings, but they also probably have parents, who are adults). However, the amount of contact participants have with children was not considered in this study, so the effect of familiarity cannot be completely ruled out.

An alternative explanation would be distinctiveness. Although distinctiveness, in the literature on face perception, has been found to increase with age (Deffenbacher et al., 1998), i.e. older faces are rated as more distinct, those studies did not include children and were from a different sensory modality (visual). Newborns' crying vocalisations, for instance, have a fundamental frequency (f0) between 400 and 600 hertz. This frequency has a rapid decrease during the first three years, to approximately 300 hertz (Robb & Saxman, 1985), and after that a slower decrease until puberty (McAllister & Sjölander, 2013). The adult's and older adult's f0 are around 130 hertz (Goy, Fernandes, Pichora-Fuller, & Van Lieshout, 2013), which is two times lower than that of children. These acoustical differences could make children's voices more distinctive, making it easier to estimate their age. Although in Amorim et al. (2019, in press) children being more accurate in the recognition of emotion in children's voices

seems to suggest an own age bias, only children evidenced a facilitated recognition of the emotion of a voice when that voice was from someone of their age. An alternative explanation could be that this is the result of children's voices being more distinctive. Instead of an own age bias, children would be better at estimating the age of other children for two reasons, (1) the age of children voices is easier to estimate, because their voices are acoustically more distinctive and (2) they do not have a lot of previous experience in estimating the age of other age groups, so they do not develop *expertise* in estimating their age. For younger adults and adults, although children's voices would also be easier to estimate, previous experience with other age groups would increase their *expertise* in estimating the age of voices of other age groups. Therefore, the differences between the accuracy in estimating the age of children speakers and other (less distinctive) age groups would be smaller. Supporting this, there were no differences in accuracy recognising the emotion of children, younger adult and older adult speakers, for adolescents and younger adult listeners (Amorim et al., 2019, in press). This is also supported by evidence of no differences in amplitude of the ERP components N170 and P2 when adults saw faces of children, adults and older adults (Melinder et al., 2010).

Because in the current study, participants seemed to have difficulties in estimating the age of different speakers, it could be that their previous experience with less distinctive age groups (such as adolescents, younger adults and adults) was not effective in increasing the accuracy in estimating the age of those age groups. In this case, because of the task difficulty, participant's familiarity with less distinctive voices, was not enough to increase accuracy in estimating their age. It could be argued that older adult's voices could also be more acoustically distinctive (from the literature of face perception Deffenbacher et al., 1998), hence their age should have been more

accurately perceived as well. However, older adult speakers are known to have physical changes in the apparatus responsible for voice production (Hollien, 1987), which makes their voices harder to perceive, especially for younger adults (Huntley et al., 1987; Neiman & Applegate, 1990; Shipp & Hollien, 1969). The accuracy in estimating the age of adult speakers, interestingly, did not differ from that of children speakers. This could be due to adult's voice having already some acoustical features that make them more distinctive and, at the same time, not having enough physical change (associated with aging) in the voice production apparatus to make their voices hard to perceive. In this task, adult speakers would be easier to estimate as their voice might already be slightly more distinctive than adolescents and younger adults (Deffenbacher et al., 1998), but probably not as distinctive as children's voice. Adult voice's age would also be easier to estimate than that of older adults, because the latter already exhibits changes (associated with aging) that make their voice hard to perceive (Hollien, 1987).

The estimation of the age of male speakers was more accurate than of female speakers. This is not surprising if we consider that male and female voices develop differently. With puberty, male and female voices become less similar and with menopause they become more similar again (Hollien, 1987). Specifically, male voices appear to have more age-related changes in some acoustical features of the voice (Hollien, 1987; McAllister & Sjölander, 2013), meaning that there are more inter-age differences, making it easier to identify the age of a male speaker. Importantly, in the current study we found an interaction between the age of the speaker and the sex of the speaker. This interaction suggests that male speakers' age is easier to estimate only in certain age groups. As these age groups correspond to those when the male and female voice are more acoustically different (i.e., younger adults, adults and older adults), this could mean that the advantage for male speakers (i.e., higher accuracy in estimating the

age of a male speaker) only holds true when there are marked acoustical changes resulting in higher sexual dimorphism of the voices, after puberty. Accordingly, humans were found to have greater f0 sexual dimorphism than any other ape (Puts et al., 2016). As female voices seem to have smaller changes in acoustical features across the lifespan (Hollien, 1987), it seems that, from an evolutionary point of view, it is more important for male voices to signal certain characteristics, such as age, probably because of its relation to other physical characteristics, such as height and masculinity, and social traits, such as dominance (Bartes et al., 2015). Supporting this, cortisol and testosterone levels interact in predicting f0 in male voices but are unrelated to f0 in female voices (Puts et al., 2016), once more suggesting that it is more important for male voices to signal certain characteristics from the voice. The fact that the advantage for male speakers only appeared after adolescence suggests that acoustical features of the voice probably evolved alongside hormonal production changes (e.g. higher production of testosterone) (Hollien, 1987). Future studies should focus on voice as an important contributor to mate selection, intra-sexual competition (Puts et al., 2016) and social hierarchy (Bartes et al. 2015; Ko, Sadler, & Galinsky, 2015). As we did not measure acoustical parameters of the voice (including distinctiveness), this is still very speculative. Future studies should consider the possibility that distinctiveness and other acoustical parameters play a role in the estimation of physical characteristics, such as age and sex.

It could be argued that because our participants were mostly females, this could result in more motivation in accurately perceiving the age of male compared to female voices, as they would be more interested in engaging with males for reproductive reasons (Kiiski et al., 2016). Firstly, in our study, sexual orientation was not considered, meaning that we do not know which sex elicits more sexual interest. Second, even if we

supposed that our participants sexual interest is directed to males, motivation to engage should probably be constrained to certain age groups, possibly matching potential romantic or sexual partners. In this study, the age of male older adult speakers was also better estimated than that of female older adult speakers and it is unlikely that older adult voices trigger more sexual interest. Supporting this, attractiveness ratings were found to decrease with age (Kiiski et al., 2016), meaning that younger adults are rated as significantly more attractive, thus, with higher reproductive value, than older adults. Hence, the interaction between speaker age and speaker sex does not seem to be driven by motivation of the listeners, but instead, by differences in perceptual cues of the voice, across age and between sex.

The interaction between speaker age and emotion category is unexpected, because the perception of identity information (such as age) and the perception of emotional cues is functionally dissociable (Belin, 2017). We found that emotion of the voice does not seem to be completely irrelevant when it comes to estimate the age of speakers. This is consistent with the existence of emotion specific age-related changes (Amorim et al., 2019, in press). Specifically, these authors found that the accuracy in the recognition of an emotion from a voice seems to vary as a function of speaker age and emotion category. In the present study, accuracy in estimating the age of vocalisations expressing enjoyment, disgust and anger did not differ as a function of speaker age. Despite this, accuracy for vocalisations expressing pleasure and neutral vocalisations varied as a function of speaker age. When the emotion expressed was pleasure, the age of children speakers was more accurately perceived than of adolescent and older adult speakers. Additionally, for neutral vocalisations, the age of children speakers was more accurately perceived than of older adult speakers. This finding is

also in line with the existence of emotion-specific age-related changes (Amorim et al., 2019, in press) in the voice.

Social Trait Overgeneralisation

The inference of the two main social traits (dominance and trustworthiness)(McAleer et al., 2014) from the voice, has shown to be the same, regardless of the sex of the evaluator. This is in line with more recent findings suggesting no differences in judgement of these social traits between male and female raters (Ponsot et al., 2018). This is important because it means that this inference is robust to variations in sex, age (Zebrowitz & Montepare, 2008 for face literature) and culture/language (Baus, McAleer, Marcoux, Belin, & Costa, 2019) of the listener.

In this study, as expected, children and adolescents were rated as less dominant and adults as more dominant. Also, as expected, male voices were rated as more dominant than female voices. Importantly, these differences between dominance ratings of male and female speakers were greater for adolescent, younger adult, adult and older adult speakers than for children speakers, as suggested by previous literature (Hollien 1987; Batres et al., 2015 for face literature). This speaker age group and speaker sex interaction was expected. As, for instance, masculinity and height are positively correlated with dominance ratings (Batres et al., 2015), voices resembling the mental representation of a masculine or tall individual are perceived as more masculine and taller and will likely be perceived also as more dominant (Zebrowitz, 2017). This is in line with the metaphorical overgeneralisation/overgeneralisation hypothesis (Secord & Stritch, 1960; Zebrowitz, 2017; Zebrowitz & Montepare, 2008).

It should be mentioned that, because in this study we know that voices of certain age groups were wrongly perceived as belonging to another age group (e.g. adolescents

and older adults are more frequently perceived as younger adults and adults, respectively), the dominance ratings of the former age group might, in fact, be those belonging to the latter age group. Specifically, the lack of differences between ratings of dominance in adult voices and older adult voices may be due to older adults speakers being perceived as adult speakers. The same can be said about adolescent's voice not differing from younger adult's voice, in dominance ratings. Despite this, our results partly replicate the inverted-U shape for the relationship between dominance ratings and chronological age found for faces (Batres et al., 2015).

Sex differences are a consistent finding and low-pitched voices (usually belonging to male speakers) have systematically been rated as more dominant by both female and male listeners(Klofstad et al., 2012; Ponsot et al., 2018; Tsantani et al., 2016). Notwithstanding, without analysing acoustic parameters, we do not know if this is a true sex effect (i.e., women are perceived as less dominant) or if it is a pitch effect (i.e., higher pitched voices are perceived as less dominant). Even in studies where acoustical parameters are taken into account, one cannot disentangle those effects as they overlap, i.e. female voices usually correspond to the high-pitched voice group and male voices usually correspond to the low-pitched voice group. This would clarify if the perception of social traits is more purely tuned to the processing of acoustic features of the voice (face literature supporting perceptually driven impressions Adams et al., 2012; Hess, Adams, Grammer, & Kleck, 2009) or if it derives from stereotypes (Hess et al., 2000). This could be achieved by comparing ratings of high-pitched male and female voices (that do not differ in pitch) and low-pitched male and female voices (that also do not differ in pitch). A sex effect would imply that female voices were perceived as less dominant and a pitch effect would imply that high pitched voices were perceived as less dominant.

Emotion and Social Trait Inference

The overgeneralisation hypothesis (Schirmer et al., 2019; face literature Zebrowitz, 2017; Zebrowitz & Montepare, 2008) used to describe the role of emotion on the inference of trustworthiness, also holds true for dominance (face literature Keltner et al., 2003). In particular, anger, signalling a more probable dominant behaviour, elicits higher ratings of dominance as a present trait in that individual (Sutton et al., 2019). So, this overgeneralisation might be the mechanism by which emotion signals the presence of a certain personality trait (support from the literature for faces - Adams et al., 2012; Hess et al., 2009; Zebrowitz, 2017; Zebrowitz & Montepare, 2008). The emotion category that drives judgements of personality traits, nevertheless, seems to be different regarding the trait being assessed. For dominance, it seems that the approach system is a possible explanation (Keltner et al., 2003). This hypothesis states the expression of certain emotions is more characteristic of individuals having high power, thus the perception of the expression of that emotion, will elicit the perception of that person as more powerful. On the contrary, the expression of a certain emotion that is more characteristic of individuals with low power, would elicit the perception of that person as having low power. Specifically, our results suggest that it is the presence of approach-related emotion (such as anger and enjoyment) that are responsible for the increase in dominance ratings. This is not in line with previous findings, which suggest that it is the inhibition-related system that decreases those ratings (Hareli et al., 2009). These differences might be due to the use of different emotion categories, i.e. in the study of Hareli et al. (2009) they used sadness and fear as inhibition-related emotion and in the current study disgust was used as an inhibition-related emotion. It could also reflect a difference in sensory modality (visual vs. auditory). One important result was

that the approach-related emotions did not equally influence dominance ratings. Anger, besides significantly differing from inhibition-related and neutral emotional expression, also stood out (significantly differed), in dominance ratings, from other approach-related emotions, such as pleasure. Pleasure, on the other hand, did not increase dominance ratings compared to neutral vocalisations or disgust. Although dominance ratings for anger and enjoyment did not differ (which is suggestive of the role of the approach-related system), differences found in ratings of voices expressing pleasure and anger, and lack of difference in ratings of voices expressing pleasure and disgust, suggests that emotions inside each system (approach and inhibition) do not linearly increase dominance ratings. This would mean that probably some emotions are more strongly associated with dominance, than others. Dividing emotion as approach-related or inhibition-related might be excessively parsimonious, and in this case, may disregard the different contribution of specific emotions to impression formation. The fact that emotion categories are differently related to dominance could also explain the inconsistency in our results and those of Hareli et al. (2009).

An alternative explanation would be that all the emotions used in this study are approach-related. Disgust, theoretically, fits the inhibition system, because it is a negative emotion, which is related to an avoidance reaction and an alarm-vigilant system (Amorim et al., 2019, in press; for face literature, Keltner et al., 2003). Nevertheless, disgust has not been studied in the context of dominance inference, hence, one cannot be entirely sure about how this emotion is expected to relate to dominance. If disgust is also an approach-related emotion, it could the case that these differences observed are negligible in comparison with differences if inhibition-related emotions were included. It would also be congruent with the majority of the emotional voices not being different from the neutral ones. It does not however, exclude that different

emotions inside each system, have different contributions to dominance ratings. Another relevant aspect is that it seems that the relation between perceived age and dominance, and chronological/real age and dominance might not be the same. This distinction is important as using different variables could lead to different results, i.e. dominance ratings varying with chronological age vs. perceived age may result in inverted-U shape relation vs. linear relation.

Regarding trustworthiness, it did not vary across age groups, which is not in line with previous studies (Schirmer et al., 2019). Younger adults did not consider their counterparts (younger adult speakers), or for that matter, any other age group, more trustworthy than the others. It may be that when it concerns general (meaning not context specific) trustworthiness, there is no age group being preferred, *a priori*. A trust preference for a specific age group may arise from a specific context. For instance, general trust was associated with economic trust but not consistently with mating-related trust (O'Connor & Barclay, 2017). Supporting this, there is an inconsistent preference for low-pitched voices, depending on trusting context and speaker sex (Montano et al., 2017). While dominance seems to be much more linearly associated with age, the same cannot be said about trustworthiness, as it seems to vary with age and context. Future studies should focus on the effect of context in the relation between trust and age. The differences between male and female voices may be, as referred previously in the case of dominance, confound of the effect of pitch.

For trustworthiness, it is the valence of the emotion that is related to the inference of this trait from a voice (Schirmer et al., 2019; from the face literature Caulfield et al., 2016; Sutton et al., 2019). Interestingly, it was not the expression of positive emotion responsible for the increase trustworthiness. It was, in fact, the expression of negative emotions that signalled an untrustworthy behaviour, hence,

untrustworthy ratings. The role of emotion in the perception of these social traits is rather delicate, because although the two dimensions (trustworthiness and dominance) are theoretically orthogonal, manipulation of faces/voices in one trait affects the perception of the other trait as well (Ponsot et al., 2018; for faces Dotsch & Todorov, 2012), which mean that they share something in common. Specifically, increasing perceived trustworthiness of a voice significantly decreased perceived dominance of that voice and decreasing perceived trustworthiness of a voice significantly increased perceived dominance of that voice (Ponsot et al., 2018). In the case of trustworthiness, voices expressing pleasure had significantly higher trustworthiness ratings that voices expressing enjoyment. This difference could be due to enjoyment being simultaneously rated as high in dominance (Hareli et al., 2009; from the face literature Keltner et al., 2003; Sutton et al., 2019). One possible explanation is that enjoyment, as a positive emotion, (1) would increase perceived trustworthiness (Schirmer et al., 2019; also Caulfield et al., 2015; Sutton et al., 2019 from the face literature) *and* (2) it would increase perceived dominance (Keltner et al., 2003; also Sutton et al., 2019 from the face literature). Therefore, trustworthiness ratings of voices expressing enjoyment, would be higher than those of a negative emotion, but the fact that it also signals dominance would have a cost on trustworthiness. This would result in trustworthiness ratings for enjoyed voices being lower than for other positive emotions that do not increase dominance. This would also explain the lack of difference in dominance ratings between enjoyment vs. pleasure, and neutral vs. disgust. Enjoyment and pleasure increase dominance (Keltner et al., 2003), but because, at the same time, they increase trustworthiness (Schirmer et al., 2019; also Caulfield et al., 2015; Sutton et al., 2019 from the face literature), they will be perceived as less dominant than those emotions which increase dominance and decrease trust (for instance, anger). This negative

correlation between the two main dimensions (Dotsch & Todorov, 2012) would explain why emotions that share properties responsible for increasing the perception of a trait, still differ on that trait's perception. Once more, this suggests that, although valence of the emotion has proven to influence perceived trustworthiness (Schirmer et al., 2019; also, Caulfield et al., 2015; Sutton et al., 2019 from the face literature), a more specific categorisation of emotions should probably be used to describe the relation between the inference of social traits and the perception of a certain emotion category

Importantly, we found no interaction between emotion category and speaker age group or speaker sex for neither dominance nor trustworthiness ratings. This suggests that the mechanisms by which emotion is related to the perception of social traits are relatively universal across age and sex (Zebrowitz, 2017; Zebrowitz & Montepare, 2008). Also, social trait inference from voice is similar between congenital blind individuals and individuals with normal vision (Oleszkiewicz, Pisanski, Lachowicz-Tabaczek, & Sorokowska, 2017), which also suggests that visual input is not necessary for this overgeneralisation to happen.

Limitations

Something worth mentioning is that the scale used for age estimation has an inconvenient consequence. A priori, children's voices can only be overestimated, and older adults' voices can only be underestimated, whereas adolescents, younger adults and adults can be both underestimated and overestimated. For the two extremes of the scale (children speakers and older adult speaker) the probability of being overestimated and underestimated, respectively, is not the same as for the age groups in between (adolescents, younger adults, adults). This means that it is easier to find a consistent pattern of results showing underestimation of age for older speakers and overestimation

of for younger speakers. Nevertheless, this would not necessarily mean that these age group voices are more underestimated or more overestimated, than the other age groups. So, instead of an underestimation or overestimation of those age groups *per se*, this would be the result of how the scale is organised. If we expanded the speaker age groups to 3-year-olds and to 80-year-olds, children (8-11 years) and older adult (60-70 years) voices could be as underestimated and overestimated as the age groups in between. In order to better assess this, it would be interesting to include response options which are outside the real age of the speaker sample (i.e. including a response option of 3-year-old, without having 3-year-old speakers).

Another important aspect is that adolescent speakers were mostly perceived as younger adult speakers and older adult speakers were mostly perceived as adult speakers. This means that, if the inference of social traits is stereotypically driven, instead of analysing the dominance/trustworthiness ratings of speaker from five age groups, we are actually only analysing ratings for speakers from three age groups (those perceived as children, those perceived as younger adults and those perceived as older adults). Although studies from faces suggest that the inference of social traits seems to be perceptually driven (Adams et al., 2012; Hess et al., 2009), this possibility cannot be completely ruled out.

In the age estimation task, the interaction between speaker age group and emotion category is delicate. Since there are many age groups and many emotion categories, the speaker's sample should be larger to prevent type I errors and ensure that the interaction is meaningful.

Future studies should include listeners from different age groups, in order to understand the role of listener age on the estimation of age and social traits of speakers also from different age groups. Describing the acoustic patterns of the different voices

and its relation to the estimation of age and social trait inference is of the utmost importance. Without it, although we can indirectly associate certain age groups with certain traits, we are blind to the mechanisms behind it and, critically, to how they developed. The understanding of psychophysical properties of sound is also necessary if we want to apply this knowledge, for instance, in manipulating voice parameters to increase the perception of certain social traits.

Conclusion and Future Concerns

Firstly, this study seems to suggest further similarities between mechanisms involved in the perception of identity information (such as age) and perception of social traits (such as dominance and trustworthiness) from the voice and previous findings from the face literature. Many known mechanisms in face perception were still unknown in voice perception. Secondly, an own age bias in estimating the age of a speaker was not found, contradicting previous studies. Instead, it seems that the age children speakers is easier to estimate, possibly due to their voices being acoustically more distinctive. All the other age groups were rather wrongly estimated, which is also not in line with previous reports of high accuracy in estimation of age from the voice. Thirdly, we replicated an inverted- U shape for the relation between dominance and age, but we found no relation between trustworthiness and age. There was also a sex effect, with male voices being perceived as more dominant and less trustworthy and female voices as less dominant and more trustworthy. Lastly, the overgeneralisation hypothesis from the face literature was also congruent for voices. Negative emotion decreased perceived trustworthiness, and approach-related emotion partly increased perceived dominance. However, this overgeneralisation seems to be emotion specific.

There are numerous situations involving voice processing mechanisms without a visual input associated, for example telephone calls and audio technology in artificial intelligence, experiencing verbal hallucinations (Ponsot et al., 2018) and congenital blindness (Oleszkiewicz et al., 2016). Also, investigation of schizophrenia and autism spectrum disorders could benefit from studies of social trait inference based on the voice, in order to understand voice processing impairments (Ponsot et al., 2018). Outside the spectrum of voice disorders, for instance, in transsexual male-to-female transition, there seems to be a relation between self and other voice perception, and quality of life (Hancock, Krissinger, & Owen, 2011). Adjusting hormonal treatment (known to affect vocal parameters) to increase identification with vocal changes could increase quality of life and adaptation to transition. Stephen Hawkins was unable to use muscles related to voice production, so he used a technological device that produced his speech. Knowledge about vocal acoustical parameters and speech prosody could contribute to help people with similar conditions to adapt to this technological equipment. Costumer service could benefit from the use of audio technology specifically designed to be perceived as competent, helpful and trustworthy, boosting client satisfaction. There are limitless applications for studies of voice perception, further knowledge would be certainly useful.

Referências Bibliográficas

Adams, R. B., Nelson, A. J., Soto, J. A., Hess, U., & Kleck, R. E. (2012). Emotion in the neutral face: A mechanism for impression formation? *Cognition and Emotion*, *26*(3), 431–441. https://doi.org/10.1080/02699931.2012.666502

Amorim, M., Anikin, A., Mendes, A.J., Lima, C., Kotz, S.A., Pinheiro, A.P. (in press). Changes in Vocal Emotion Recognition Across the Life Span. Emotion.

Anastasi, J. S., & Rhodes, M. G. (2005). An own-age bias in face recognition for children and older adults. *Psychonomic Bulletin and Review*, *12*(6), 1043–1047.

Audacity Team (2019). Audacity(R): Free Audio Editor and Recorder [Computer application]. Version 2.3.2 retrieved May 20th 2019

Batres, C., Re, D. E., & Perrett, D. I. (2015). Influence of Perceived Height, Masculinity, and Age on Each Other and on Perceptions of Dominance in Male Faces. *Perception*, *44*(11), 1293–1309. https://doi.org/10.1177/0301006615596898

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

Belin, P. (2017). Similarities in face and voice cerebral processing. *Visual Cognition*, *25*(4–6), 658–665. https://doi.org/10.1080/13506285.2017.1339156

Belin, P., Fecteau, S., & Bédard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135. https://doi.org/10.1016/j.tics.2004.01.008

Boersma, Paul & Weenink, David (2019). Praat: doing phonetics by computer [Computer program]. Version 6.1.04

Bruyer, R., Mejias, S., & Doublet, S. (2007). Effect of face familiarity on age decision. *Acta Psychologica*, *124*(2), 159–176. https://doi.org/10.1016/j.actpsy.2006.03.001

Carstensen, L. L. (1995). Evidence for a Life-Span Theory of Socioemotional Selectivity. *American Psychological Society*, *4*(5), 152–156.

Caulfield, F., Ewing, L., Bank, S., & Rhodes, G. (2016). Judging trustworthiness from faces: Emotion cues modulate trustworthiness judgments in young children. *British Journal of Psychology (London, England : 1953)*, *107*(3), 503–518. https://doi.org/10.1111/bjop.12156

Deffenbacher, K. A., Vetter, T., Johanson, J., & Otoole, A. J. (1998). Facial aging, attractiveness, and distinctiveness. *Perception*, *27*, 1233–1243.

Dotsch, R., & Todorov, A. (2012). Reverse Correlating Social Face Perception. *Social Psychological and Personality Science*, *3*(5), 562–571. https://doi.org/10.1177/1948550611430272

Goy, H., Fernandes, D. N., Pichora-Fuller, M. K., & Van Lieshout, P. (2013). Normative voice data for younger and older adults. *Journal of Voice*, *27*(5), 545–555. https://doi.org/10.1016/j.jvoice.2013.03.002

Goy, H., Kathleen Pichora-Fuller, M., & van Lieshout, P. (2016). Effects of age on speech and voice quality ratings. *The Journal of the Acoustical Society of America*, *139*(4), 1648–1659. https://doi.org/10.1121/1.4945094

Hancock, A. B., Krissinger, J., & Owen, K. (2011). Voice perceptions and quality of

life of transgender people. *Journal of Voice*, *25*(5), 553–558. https://doi.org/10.1016/j.jvoice.2010.07.013

Hareli, S., Shomrat, N., & Hess, U. (2009). Emotional Versus Neutral Expressions and Perceptions of Social Dominance and Submissiveness. *Emotion*, *9*(3), 378–384. https://doi.org/10.1037/a0015958

Hess, U., Adams, R. B., Grammer, K., & Kleck, R. E. (2009). Face gender and emotion expression: Are angry women more like men? *Journal of Vision, 9*(12), 19–19. https://doi.org/10.1167/9.12.19

Hess, U., Senecal, S., Kirouac, G., Herrera, P., Philippot, P., & Kleck, R. E. (2000). Emotional expressivity in men and women: Stereotypes and self-perceptions. *Cognition and Emotion*, *14*(5), 609–642. https://doi.org/10.1080/02699930050117648

Hills, P. J., & Lewis, M. B. (2011). The own-age face recognition bias in children and adults. *Quarterly Journal of Experimental Psychology*, *64*(1), 17–23. https://doi.org/10.1080/17470218.2010.537926

Hollien, H. (1987). “Old Voices”: What Do We Really Know About Them? *Journal of Voice*, *1*(1), 2–17.

Huntley, R., Hollien, H., & Shipp, T. (1987). *Influences of Listener Characteristics on Perceived Age Estimations*. *1*(1), 49–52.

Keltner, D., Gruenfeld, D. H., & Anderson, C. (2003, April). Power, Approach, and Inhibition. *Psychological Review*, Vol. 110, pp. 265–284. https://doi.org/10.1037/0033-295X.110.2.265

Kiiski, H. S. M., Cullen, B., Clavin, S. L., & Newell, F. N. (2016). Perceptual and social attributes underlining age-related preferences for faces. *Frontiers in Human Neuroscience*, *10*(AUG2016). https://doi.org/10.3389/fnhum.2016.00437

Klofstad, C. A., Anderson, R. C., & Peters, S. (2012). Sounds like a winner: Voice pitch influences perception of leadership capacity in both men and women. *Proceedings of the Royal Society B: Biological Sciences*, *279*(1738), 2698–2704. https://doi.org/10.1098/rspb.2012.0311

Ko, S. J., Sadler, M. S., & Galinsky, A. D. (2015). The Sound of Power: Conveying and Detecting Hierarchical Rank Through Voice. *Psychological Science*, *26*(1), 3–14. https://doi.org/10.1177/0956797614553009

Labouvie-Vief, G. (2003). Dynamic Integration: Affect, Cognition, and the Self Adulthood. *Memory & Cognition*, *12*(6), 202–206.

Li, Q., Heyman, G. D., Mei, J., & Lee, K. (2019). Judging a Book by Its Cover: Children’s Facial Trustworthiness as Judged by Strangers Predicts Their Real-World Trustworthiness and Peer Relationships. *Child Development*, *90*(2), 562–575. https://doi.org/10.1111/cdev.12907

Lima, C. F., Castro, S. L., & Scott, S. K. (2013). When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods*, *45*(4), 1234–1245. https://doi.org/10.3758/s13428-013-0324-3

Little, A. C. (2012). Manipulation of infant-like traits affects perceived cuteness of infant, adult and cat faces. *Ethology*, *118*(8), 775–782. https://doi.org/10.1111/j.1439-0310.2012.02068.x

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say “hello”? Personality

impressions from brief novel voices. *PLoS ONE*, *9*(3). https://doi.org/10.1371/journal.pone.0090779

McAllister, A., & Sjölander, P. (2013). Children's voice and voice disorders. *Seminars in Speech and Language*, *34*(2), 71–79. https://doi.org/10.1055/s-0033-1342978

Melinder, A., Gredebäck, G., Westerlund, A., & Nelson, C. A. (2010). Brain activation during upright and inverted encoding of own- and other-age faces: ERP evidence for an own-age bias. *Developmental Science*, *13*(4), 588–598. https://doi.org/10.1111/j.1467-7687.2009.00910.x

Michalsky, J., & Schoormann, H. (2017). Pitch convergence as an effect of perceived attractiveness and likability. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, *2017-August*, 2253–2256. https://doi.org/10.21437/Interspeech.2017-1520

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

Nations, U., of Economic, D., Affairs, S., & Division, P. (2019). *World Population Prospects 2019 Highlights*. (ST/ESA/SER.A/423).

Neiman, G. S., & Applegate, J. A. (1990). Accuracy of Listener Judgements of Perceived Age Relative to Chronological Age in Adults. *Folia Phoniatr*, *42*, 327–330.

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

O'Connor, J. J. M., & Barclay, P. (2018). High voice pitch mitigates the aversiveness of antisocial cues in men's speech. *British Journal of Psychology*, *109*(4), 812–829. https://doi.org/10.1111/bjop.12310

Oleszkiewicz, A., Pisanski, K., Lachowicz-Tabaczek, K., & Sorokowska, A. (2017). Voice-based assessments of trustworthiness, competence, and warmth in blind and sighted adults. *Psychonomic Bulletin and Review*, *24*(3), 856–862. https://doi.org/10.3758/s13423-016-1146-y

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

Psychology Software Tools, Inc. [E-Prime 2.0]. (2012). Retrieved from https://www.pstnet.com.

Puts, D. A., Hill, A. K., Bailey, D. H., Walker, R. S., Rendall, D., Wheatley, J. R., … Ramos-Fernandez, G. (2016). Sexual selection on male vocal fundamental frequency in humans and other anthropoids. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1829). https://doi.org/10.1098/rspb.2015.2830

Robb, M. P., & Saxman, J. H. (1985). Developmental Trends in Vocal Fundamental Frequency of Young Children. *Journal of Speech, Language, and Hearing Research*, *28*(3), 421–427. https://doi.org/10.1044/jshr.2803.427

Rossion, B. (2014). Understanding face perception by means of prosopagnosia and neuroimaging. *Frontiers in Bioscience*, *6*, 258–307.

Schirmer, A., Feng, Y., Sen, A., & Penney, T. B. (2019). Angry, old, male – and

trustworthy? How expressive and person voice characteristics shape listener trust. *PLoS ONE, 14*(1). https://doi.org/10.1371/journal.pone.0210555

Schroder, M. (2003). Experimental study of affect bursts. *Speech Communication, 40*, 99–116.

Secord, P. F., & Stritch, T. F. (1960). The Role of Metaphorical Generalization and Congruency in the Perception of Facial Characteristics. *Journal of Social Psychology, 52*(2), 329–337. https://doi.org/10.1080/00224545.1960.9922088

Sen, A., Isaacowitz, D., & Schirmer, A. (2018). Age differences in vocal emotion perception: on the role of speaker age and listener sex. *Cognition and Emotion, 32*(6), 1189–1204. https://doi.org/10.1080/02699931.2017.1393399

Shipp, T., & Hollien, H. (1969). Perception of the Aging Male Voice. *Journal of Speech and Hearing Research, 12*(4), 703–710. https://doi.org/10.1044/jshr.1204.703

Shipp, T., Qi, Y., Huntley, R., & Hollien, H. (1992). Acoustic and temporal correlates of perceived age. *Journal of Voice, 6*(3), 211–216. https://doi.org/10.1016/S0892-1997(05)80145-6

Stahl, J., Wiese, H., & Schweinberger, S. R. (2008). Expertise and own-race bias in face processing: an event-related potential study. *Cognitive Neuroscience and Neuropsychology, 19*(5), 583–587.

Sutton, T. M., Herbert, A. M., & Clark, D. Q. (2019). Valence, arousal, and dominance ratings for facial stimuli. *Quarterly Journal of Experimental Psychology, 72*(8), 2046–2055. https://doi.org/10.1177/1747021819829012

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., Said, C. P., Engell, A. D., & Oosterhof, N. N. (2008). Understanding evaluation of faces on social dimensions. *Trends in Cognitive Sciences*, *12*(12), 455–460. https://doi.org/10.1016/j.tics.2008.10.001

Tsantani, M. S., Belin, P., Paterson, H. M., & McAleer, P. (2016). Low Vocal Pitch Preference Drives First Impressions Irrespective of Context in Male Voices but Not in Female Voices. *Perception*, *45*(8), 946–963. https://doi.org/10.1177/0301006616643675

Van Lancker, D. R., Cummings, J. L., Kreiman, J., & Dobkin, B. H. (1988). Phonagnosia: A Dissociation Between Familiar and Unfamiliar Voices. *Cortex*, *24*(2), 195–209. https://doi.org/10.1016/S0010-9452(88)80029-7

van Rijsbergen, N., Jaworska, K., Rousselet, G. A., & Schyns, P. G. (2014). With age comes representational wisdom in social signals. *Current Biology : CB*, *24*(23), 2792–2796. https://doi.org/10.1016/j.cub.2014.09.075

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.

Zebrowitz, L. A. (2017). First impressions from faces. *Current Direction of of Psychological Sciences*, *26*(3), 237–242.

Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Social and Personality Psychology Compass*, *2*(3), 1497–1517. https://doi.org/10.1111/j.1751-9004.2008.00109.x

Appendix A

Participants' agreement on estimation of age, dominance ratings and trustworthiness ratings

	N	ICC	Sig.	Cronbach's alfa
Estimation of age	28	0.981	<0.001	0.982
Dominance ratings	30	0.886	<0.001	0.917
Trustworthiness ratings	29	0.860	<0.001	0.891

Note. N = number of participants; ICC = Intraclass Correlation Coeficient; ICC is estimated for mean absolute agreement between answers.

Appendix B

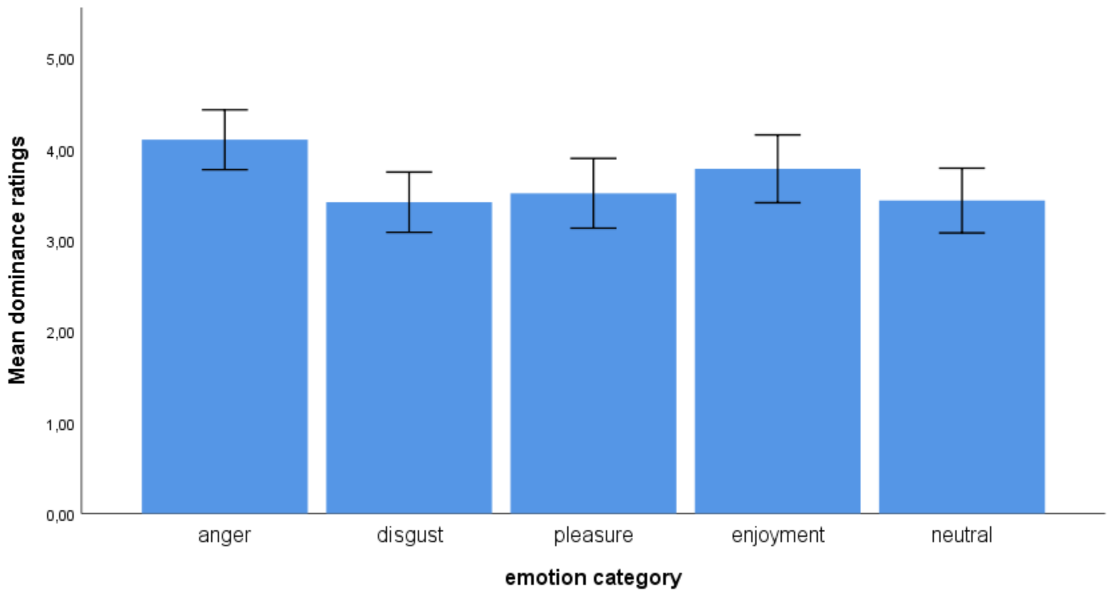
Mean ratings (and SD) of dominance and trustworthiness for each age group and sex

		Dominance Ratings		Trustworthiness Ratings	
		Mean	SD	Mean	SD
Children	Female	2.832	0.535	3.889	0.196
	Male	3.086	0.213	3.822	0.265
Adolescents	Female	3.046	0.120	3.722	0.132
	Male	4.132	0.192	3.715	0.173
Younger Adults	Female	3.107	0.152	4.119	0.223
	Male	4.207	0.162	3.685	0.167
Adults	Female	3.761	0.162	3.581	0.260
	Male	4.332	0.196	3.722	0.260
Older Adults	Female	3.646	0.205	3.889	0.207
	Male	4.364	0.219	3.278	0.278

Note. Dominance and trustworthiness ratings ranged from 1 to 7.

Appendix C

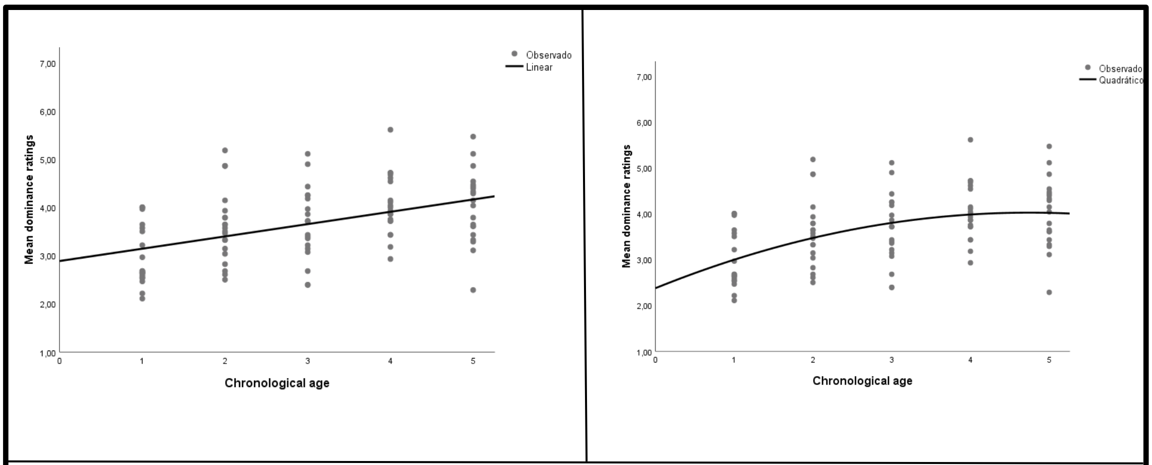
Mean Dominance Ratings for each Emotion



Bars represent 95% Confidence Interval (CI). Voices expressing anger were perceived as significantly more dominant than those expressing disgust (p<0.001), pleasure (p=0.003) and neutral ones (p<0.001), but not enjoyment (p=0.384). Vocalisations expressing enjoyment did not differ from any other emotional category (p>0.05 for anger, disgust, pleasure and neutral) and the comparisons between the other emotional categories were also not significant (p>0.05 for all comparisons).

Appendix D

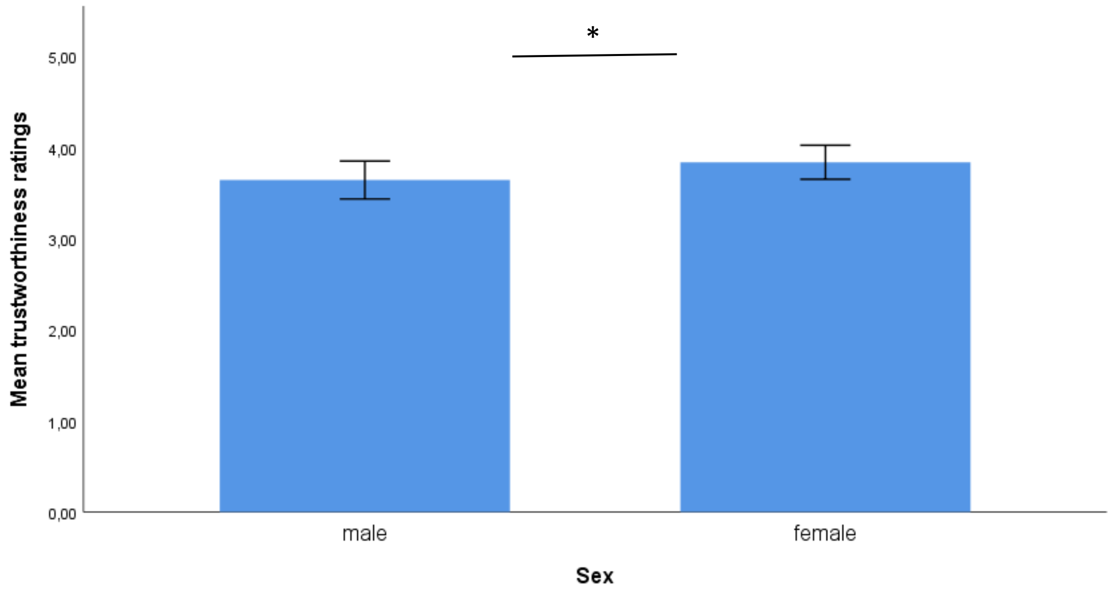
Comparison Between the Linear and Quadratic Models, for the Relation Between Dominance Ratings and Chronological Age



R^2=0.20 for the linear model (left) ; R^2=0.22 for the quadratic model (right). Chronological age ranged from 1 to 5 and dominance ratings ranged from 1 to 7.

Appendix E

Trustworthiness Ratings for each Sex



* *p*<0.05. (male speakers: Mean = 3.644, SE = 0.104; female speakers: Mean = 3.84, SE = 0.093). Trustworthiness ratings ranged from 1 to 7.

Appendix F

Mean Ratings (and SD) of Dominance and Trustworthiness for each Emotion

	Dominance Ratings		Trustworthiness Ratings	
	Mean	SD	Mean	SD
Anger	4.104	0.157	3.230	0.087
Disgust	3.418	0.159	3.148	0.098
Pleasure	3.516	0.183	4.300	0.074
Enjoyment	3.784	0.178	3.852	0.203
Neutral	3.436	0.170	4.182	0.061

Note. Dominance and trustworthiness ratings ranged from 1 to 7.