Thanh Lan Truong & Andrea Weber

2022

Trust Issues: The Effect of Speaker Age on Credibility
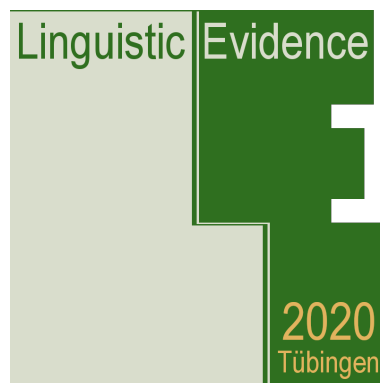
in

Trust Issues: The Effect of Speaker Age on Credibility

Thanh Lan Truong & Andrea Weber

University of Tübingen

thanh-lan.truong@uni-tuebingen.de; andrea.weber@uni-tuebingen.de

1   Introduction

Throughout our lives, how we talk and sound affects how we are perceived and judged by others. That is, whenever we speak, we are being evaluated, and the credibility of what we say is being weighted (Ferguson & Zayas, 2009). Importantly, credibility not only depends on *what* we say but also on *how* we say it. The *how* includes, for example, the nativeness of our pronunciation, such that trivia statements made by non-native speakers have been rated as less true than the same statements made by native speakers (Lev-Ari & Keysar, 2010). The present study set out to test if similar effects on credibility can be found for speakers varying in age rather than in nativeness. Specifically, trivia statements spoken by children will be compared to the same statements spoken by adults. Investigating age of speaker effects will help to understand the underlying mechanisms better that govern credibility ratings.

People with non-native accents often have to face stigmatization, social ostracism, or unfair jurisdiction (e.g., Dixon et al., 2002). It is generally acknowledged in sociolinguistics that people with a foreign accent are commonly judged as inferior (Edwards, 1999; Gluszek & Dovidio, 2010; Munro et al., 2006), in terms of intelligence, educational background, prestige, kindness, attractiveness, and trustworthiness (Anderson et al., 2007; Fuertes et al., 2002; Lev-Ari & Keysar, 2010; Lindemann, 2003). As a consequence, eyewitnesses with foreign accents are perceived as less credible than those with native accents (Frumkin & Stone, 2020). Judges, however, perceive children as more honest than adult witnesses, despite their limited memory capacities and verbal skills, which make them appear less reliable than adults (Bala et al., 2015). Unlike towards foreign-accented speakers, listeners typically have a positive attitude towards children, although the reliability of what they say might be seen to be lower than that of adult speakers.

For the effects of non-nativeness on credibility, at least two explanations have been proposed. Negative attitudes towards non-native speakers are possibly promoted by in-group biases and not by the accent as such, which serves just as the marker for the biases (e.g., Dixon et al., 2002). Alternatively, credibility depends on how easily listeners can process the linguistic signal that deviates from the norms of the target language (i.e., foreign-accented speech; (Lev-Ari & Keysar, 2010)). Given that non-native speech typically deviates from the standard norms, it can conceivably hamper *processing fluency*, which in turn could have a potential impact on listeners' credibility judgements (Oppenheimer, 2008; Unkelbach, 2006). The term *processing fluency* can be broadly described as the ease of stimulus processing. For example, if speech is easy to understand, it is perceived as not only more pleasurable (Reber et al., 2004), familiar (Whittlesea et al., 1990), and less risky (Song & Schwarz, 2009) but also as more truthful (Reber & Schwarz, 1999). For example, rhyming language is known to be easier to process, and indeed it has been found that although the phrase *woes unite enemies* has the same meaning as in *woes unite foes*, the latter is perceived as more accurate because of the rhyming of the words (McGlone & Tofighbakhsh, 2000). Children's speech is interesting in this regard since, as mentioned above, attitudes towards children are likely to be more positive than towards non-native speakers, but both varieties of speech deviate from the norms of the target language and hence

are more difficult to process.

Lev-Ari & Keysar (2010) argued in their study for processing difficulties as the driving factor for more negative credibility ratings in non-native speech. They tested three types of accents with different degrees of accent strength (native accents: English, mild non-native accent: Polish, Turkish, and Austrian-German, and heavy non-native accents: Korean, Turkish, and Italian). Native English listeners were asked to judge the veracity of each trivia statement like *ants don't sleep* on a 14 cm line with the left pole marked with *definitely false* and the right pole marked with *definitely true*. In an attempt to control negative stereotypical biases towards non-native speakers, participants were told that the non-native speakers solely acted as messengers of the statements, reciting statements which were provided by the experimenter. Thus, the statements could not reflect the speaker's educational background for example. The results showed that native listeners judged trivia statements less often as true when the statements were spoken by a foreign-accented speaker than when the speaker was native. The authors concluded from this that not so much prejudice but rather segmental and prosodic deviations from the standard norms of the target language had a negative impact on *processing fluency* (Munro & Derwing, 1995), which in turn impacted listeners' credibility judgements. Their findings propelled further researchers to investigate credibility judgements from different perspectives and in different language contexts with mixed results (e.g., Meo, 2012; Souza & Markman, 2013; Stocker, 2017). For example, Souza & Markman (2013) attempted to replicate Lev-Ari & Keysar (2010) findings, but they also tested further manipulations of processing difficulty. In the first experiment, white noise at distinct Signal-to-Noise ratios and speech babble noise were added to the recordings of female native English speakers. Neither white noise nor speech babble noise affected the evaluation of credibility, meaning that the presence of white noise in speech did not influence credibility judgements negatively. In the second experiment, foreign-accented speech was implemented, as a direct comparison to Lev-Ari & Keysar (2010). Again, they found no effect of nativeness on credibility judgements. The primary focus of Souza & Markman (2013) was set on truth judgements of foreign-accented speech by native listeners (see also, Meo, 2012; Stocker, 2017). But what about other types of deviations from the standard norms of native adult speech, like for example children's speech?

1.1 The current study

The present work rather aims at investigating the effects of speaker age on credibility judgements, using a similar methodology to the study by Lev-Ari & Keysar (2010). Similar to foreign-accented speech, acoustic and linguistic properties of children's speech are distinct from those of native adult speech (Lee et al., 1999). Generally speaking, there can be differences between adult and child speakers based on deviations in pronunciation from the adult norm as well as differences that are caused by distinct physical characteristics between the two groups of speakers. While the average vocal folds of the adult are about 12-21 mm long, the vocal folds of an 8-year-old have grown to approximately 8 mm in length. The fundamental frequency (i.e., F0) of an infant's voice is at birth by around 500 Hz. As the larynx grows with age, F0 drops to about 275 Hz by the age of eight, with little difference between boys and girls. More specifically, boys have typically lower formant frequencies than girls (Vorperian & Kent, 2007). While F0 remains quite stable throughout childhood, about 2.5 octaves, the variability decreases progressively until the age of 8-12 years (Kent, 1976). Not surprisingly, speech performance thus becomes more adult-like as children grow older. The present study concentrates on children's speech at the age of seven. Although pronunciation norms typically approximate adult performance by seven years of age, it is likely that they do not fully align with that of adult speakers yet. Furthermore, children's speech is generally characterized by greater acoustic-phonetic variation than native adult speech (Smith et al., 1983; Tingley & Allen, 1975) which might consequently impact processing fluency negatively, too. For example, children's speech displays "higher pitch and formant frequencies, longer segmental durations, and greater

temporal and spectral variability" (Lee et al., 1999: p. 1455). The principle question at issue here concerns thus whether credibility judgements are being affected by speaker age.

By investigating age of the speaker effects in the present study, we wanted to get more insights about the fundamental mechanisms of speaker-related effects on credibility. As already pointed out, the approach of the present experiment adopted the methodology from Lev-Ari & Keysar (2010) but compared truth judgements for trivia statements spoken by a native female child speaker and a native male adult speaker. Given that child speech is generally characterized by greater-acoustic-phonetic variation than is adult speech (e.g., Lee et al., 1999), which potentially causes processing difficulties, we expect lower credibility ratings for statements spoken by a child speaker than by an adult speaker.

2   Method

2.1   Material

2.1.1 Speech material

Forty-five trivia statements were taken from Lev-Ari & Keysar (2010) and translated from English into German. The majority of the trivia statements were about the animal kingdom (e.g., *Ameisen schlafen nicht*, 'ants don't sleep'). Nine of the statements were replaced by new statements because their German translation did not work well (e.g., 'The original name for butterfly was flutterby' was replaced since the German word for butterfly, *Schmetterling*, does not entail the embedded words *flutter* and *by*). The selected trivia statements were statements for which the correct answer is typically not known; thus judgements on a scale were not likely to be fixed on the endpoints of a credibility scale. This was necessary to allow differences in judgement to emerge when the same statements were produced by different speakers. In order to have occasional trials in which participants could be sure of the truth value, 14 filler statements were added (e.g., *Brokkoli ist ungesund*, 'broccoli is unhealthy'). Two experimental lists with the 45 statements and 15 filler statements were created. Each experimental sentence appeared once in each list, half of which were true and half of which were false, counterbalanced for the age of the speaker. The order of sentence presentations and speaker was pseudo-randomized. Each trial began with two practice sentences.

2.1.2 Speaker selection and recording

The sixty trivia statements were recorded by one male adult speaker of German (age 54) and one female child speaker of German (age 7), both living in Tübingen. Both speakers had no reported speech impediment. Before the recording session, the speakers had time to get acquainted with the list of statements in order to prevent any disfluencies when reading. The recording session took place in a sound-attenuated room with a high-quality microphone and a sampling rate of 44 kHz. Both speakers were recorded separately. The child speaker, however, was recorded together with her mother. While the mother read from orthographic transcription, the child was prompted to repeat after her reading.[1] Special care was taken that all sentences were produced as intended and without disfluencies. The adult speaker had an average F0 of 224.14 Hz and the child speaker had an average F0 of 287.98 Hz. The difference in F0 between these two speakers was significant ($t(118) = -2.51$, $p < .02$). The credibility of all statements was not shared with the speakers, so that their speech was not affected. The main purpose of the study was revealed after the recording session.

---

[1] Elementary school starts at the age of six in Germany, and by the age of seven reading aloud is typically still less fluent than in adult. Repeating after an auditory prompt ensured that the chid speaker produced the sentences naturally and fluently.

2.2   Participants

Forty-two native listeners of German (23 female), between 19 and 34 years old (mean age = 23.5, SD = 3.8), participated in the experiment for monetary compensation. All participants were students at the University of Tübingen. None of them suffered from any hearing disorders, and they all had intact or corrected vision.

2.3   Ethics statement

The procedures for the present experiment were approved by the DGFS (Deutsche Gesellschaft für Sprachwissenschaft) ethics committee for the Chair of Psycholinguistics and Applied Linguistics at the University of Tübingen.

2.4   Procedure

The experiment was controlled with Excel Visual Basic (version 16.0.11328.20362). Participants wore over-ear headphones and were tested individually. Instructions were presented on the computer screen. The experiment started with two practice sentences, followed by the 60 trivia statements. Each statement was presented once. Participants entered their truth judgements with the use of a sliding scale. The sliding scale was similar to the one used by Lev-Ari & Keysar (2010); the left end of the scale was labelled with *definitely false*, and the right end was labelled with *definitely true*. Participants evaluated the level of credibility for each sentence by dragging the slider bar, starting from its default position at the scale's center, until it reached the desired answer position. Although not visible to participants, the positions on the scale ranged underlyingly from 0 to 14, with higher numbers indicating higher perceived credibility. Furthermore, participants were asked to use the middle of the scale as little as possible and to genuinely try to assess the veracity of each statement. In addition, participants were asked to indicate whether they had heard this statement before and already knew the correct answer. Three possible answer options were given: *yes, no,* and *unsure*. After the experiment was completed, participants filled in a short language background questionnaire.

3   Results and discussion

R (R Core Team, 2013: version 1.2.5019) and lme4 (Bates et al., 2015) were used to perform linear mixed effects analyses on listeners' perceived truth judgements. Only statements that were unknown to the participant were analysed (73.3 % of the items), since we expected that known statements were less likely to be affected by speaker attributes,[2] and the pattern of results did not significantly change. Two participants had to be excluded since they did not follow the instructions.

---

[2] Note that Lev-Ari & Keysar (2010) had not found any evidence for an effect of *knowledge*. Identical to their statistical analysis, we additionally tested all statements and included the interaction of *age of speaker* (i.e., adult, child) and *knowledge* (i.e., yes, no, unsure) to the model. Similarly to Lev-Ari & Keysar (2010), *knowledge* did not improve the model.
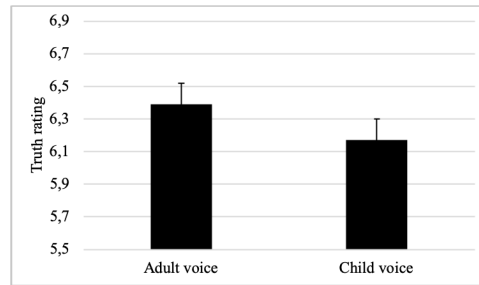
Figure 1. Figure 1. Truth ratings as a function of age of speaker voice (adult, child). The y-axis indicates the credibility ratings from *definitely false* to *definitely true*. Higher numbers indicate higher perceived credibility.

The initial model included *speaker age* (adult, child) and *participants* and *items* as random variables with random slopes. Although descriptively, ratings for the child speaker were somewhat lower than for the adult speaker, the analysis showed only a non-significant trend of *speaker age* (β = -0.06, SE = 0.04, t = -1.33, p > .1), suggesting that overall credibility ratings were not significantly affected by *speaker age*. When we further looked at the data descriptively, we noted, however, that participants varied in their response patterns. Specifically, female participants displayed different pattern from male participants. Based on this observation, we decided to conduct an exploratory analysis. We now grouped the data based on the gender of the participants (i.e., male and female listeners).
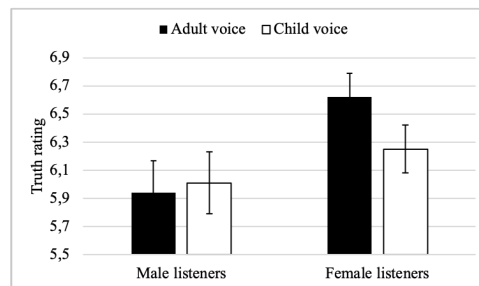

Figure 2. Figure 2. Truth ratings of male and female listeners for adult and child voices. The y-axis indicates the truth ratings from *definitely false* to *definitely true*. Higher numbers indicate higher perceived credibility.

Three participants had to be excluded from the analysis since they did not indicate their gender in the questionnaire. And indeed, when gender of the participants was considered in the analysis, there was a marginally significant effect of *age of speaker* (β = -0.11, SE = 0.06, t = - 1.82, p < .07) for female participants.[3] For male listeners, no indication of a voice effect was found (β = -0.06, SE = 0.05, t = -1.20, p > .2). The direction of the effect for female listeners was such that statements made by the child speaker were judged to be less trustworthy.

Thus, while our initial hypothesis that credibility ratings are generally affected by the age of the speaker was not confirmed, when taking participants' gender into account in an exploratory analysis, different patterns of ratings were found for female and male listeners. Specifically, female listeners judged sentences spoken by the male adult speaker as more credible than sentences spoken by the female child speaker.

---

[3] We refrained from a Bonferroni correction, as this might lead to committing a Type II error (i.e., false negative) (Winter, 2019). However, if we were to compute a pairwise comparison, using the *emmeans* (Lenth, 2018), the *p*-value would change from <.067 to < .068.

4 General discussion

The present study, motivated by Lev-Ari & Keysar (2010) findings, examined whether or not credibility judgements are being affected by speaker age. Lev-Ari & Keysar (2010) found that native listeners judged trivia statements like *ants don't sleep* as less true when the statements were spoken by a foreign-accented speaker than when the speaker was native. It was argued that this effect emerged not because of the negative prejudices towards non-native speakers, but rather because of linguistic deviations from native speech. Such deviations can conceivably hamper the ease of speech processing (i.e., "processing fluency"), which consequently can affect listeners' credibility judgements. The current study adopted the methodology from Lev-Ari & Keysar (2010) but compared truth judgements of trivia statements spoken by a male adult speaker and a child speaker. Analogous to foreign-accented speech, children's speech has been reported to deviate from native adult speech in terms of acoustic and linguistic properties (Lee et al., 1999). This variation could make speech processing conceivably more difficult, which in turn could influence credibility ratings.

The overall results showed no significant difference in truth judgements between the adult and child speaker. Thus, we failed to confirm our initial hypothesis that credibility ratings for the adult speaker will be higher than for the child speaker. As adults, we have accrued extensive knowledge of the world through experience. We expected that this experience would make it likely for everyone to trust the credibility of statements more when spoken by an adult speaker than by a child speaker. However, this is not what we found. As already mentioned in the introduction, some recent studies failed to find evidence for Lev-Ari & Keysar (2010) assumption that processing difficulties are the underlying reason for speaker effects on credibility ratings in the Italian context (Meo, 2012), in the Swiss context (Stocker, 2017), and for speech with white noise at different Signal-to-Noise Ratio and speech babble noise (Souza & Markman, 2013). The failed attempts to replicate Lev-Ari & Keysar (2010) illustrate that the effect does not easily generalize. For example, Stocker (2017) used the same set of trivia statements as in the original study and examined truth ratings but in the Swiss context. Although the methodology was relatively similar to Lev-Ari & Keysar (2010), results showed no effect of accent condition (i.e., French and German) on the evaluation of credibility. Aside from investigating the effect of foreign-accented speech on perceived credibility, Meo (2012) examined if segmental and suprasegmental acoustic features, which are typical for beginners and intermediate L2 learners of Italian, influenced credibility ratings. Such features are, for example, inadequate pauses, incorrect tonal alterations, atypical durations, slower speech rate, and disfluencies. Lower language skills increase higher percentage of speech anomalies, and consequently lower comprehensibility. Recordings of native and non-native speakers of Italian were artificially modified. Meo's (2012) findings revealed that comprehensibility mediates the relationship between foreign accent and credibility, irrespective of the speaker's accent (i.e., native accent, mild non-native accent, and strong non-native accent). Since different accents in different language contexts were tested, it is to be highlighted that the aforementioned studies were no full replication of Lev-Ari & Keysar (2010), possibly explaining the differences in results. Attitudes may vary for different accents in different countries, meaning that they are not equally negative, because accents may also evoke positive reactions in listeners' perception (Dewaele & McCloskey, 2015). For example, while the Spanish accent is considered a non-standard accent [4] in the United States, it has been shown to positively affect listener's perception of speaker's educational background, social status, and personal traits like attractiveness in the United Kingdom (Fuertes et al., 2012). Moreover, Giles (1970) found that French-accented English received more positive evaluations than Italian or German accents, even superior to English regional accents such as the Birmingham accent. This clearly illustrates that it is decidedly more complex to replicate a study where different accents are involved. A further major concern is that Lev-Ari & Keysar's (2010) approach to completely exclude stereotypes

---

[4] Non-standard accents are foreign accents spoken by a minority or lower socioeconomic group.

against non-native speakers might not have been successful and their results might have been triggered by negative attitudes towards L2 speakers after all. If stereotypes indeed play a role in credibility ratings, then our initial finding might suggest that child's speech negatively affected processing fluency and therefore lowered credibility ratings but since children are generally likeable, positive attitudes might have simultaneously increased the ratings, despite the fact that children are less knowledgeable than adults. Thus, two factors might have cancelled each other out, possibly explaining the initial null effect. However, the exploratory analysis, albeit only marginally significant, showed a different pattern when taking listeners' gender into account. For male listeners, no difference in truth judgements was obtained between the adult and child speaker, but for female listeners, truth judgements were lower for the child speaker than for the male adult speaker.[5] At first glance, this finding for female listeners appears to be in line with Lev-Ari & Keysar (2010), reasoning for non-native speakers. Recall that the effect found in Lev-Ari & Keysar (2010) was argued to be driven by the deviations in pronunciation in non-native speech which made processing harder. However, the exploratory results rather point into the direction that potentially attitudes play a role in credibility judgements, because speech intelligibility of the child speaker was not further manipulated in this study. The child speaker had produced fluent sentences without obvious mispronunciations. Also, none of the participants reported any comprehension difficulties when listening to both adult and child speaker. Therefore, processing fluency appears not to be the main reason for this finding. Meo (2012), for example, argued that poor comprehensibility as such, regardless of the strength of the foreign accent, did not influence the perceived credibility of statements but that prosodic characteristics of an utterance had a greater influence on credibility judgements. In fact, several empirical studies have demonstrated that pitch, a key acoustical feature otherwise known as fundamental frequency (i.e., F0: 'highness' or 'lowness') and its corresponding harmonics (Fitch, 2000), influence judgements of people's personality traits (Belin et al., 2011; McAleer et al., 2014; Tsantani et al., 2016).

A growing literature has demonstrated that the speaker's first impression is inferred not only from visual cues but also from auditory cues (Rezlescu et al., 2015; Zuckerman & Miyake, 1993). Listeners particularly rely on auditory cues when it comes to identifying dominance and trustworthiness impressions (Rezlescu et al., 2015). More specifically, individuals with lower voices, are perceived as taller (Xu et al., 2013), physically stronger (Sell et al., 2010), socially more dominant (Tigue et al., 2012), and more attractive (Feinberg et al., 2005). It has long been known that attractiveness is typically associated with honesty, commonly related to as the 'what is beautiful is good' stereotype (Dion et al., 1972). To put it simply, attractive voices are characterized with positive attributes such as intelligence, kindness, and trustworthiness (Hughes & Miller, 2015; Zuckerman & Miyake, 1993). Therefore, voices which are preferred are perceived as more attractive and thus more trustworthy. And indeed, the exploratory analysis showed that female listeners found sentences spoken by the male adult speaker as more credible than when spoken by the child speaker. However, it was not exclusively pitch that influenced the credibility ratings in the exploratory analysis. If the results were mainly caused by lower-pitch preference, then the pattern should be identical for female adult speakers since they have lower pitch than children.

An ongoing study by Truong & Weber (in prep.) replicated the present work, using the same trivia sentences but substituted the male speaker with a female adult speaker.[6] Contrary to the present findings, female listeners judged sentences produced by the female adult speaker as less credible compared to sentences spoken by the female child speaker. There was the possibility that this effect emerged due to a single-speaker effect. Hence, Truong & Weber (in prep.) also reproduced this experiment with multiple female adult speakers and multiple female child

---

[5] We only found a mild voice-pitch preference effect (p < .07), possibly due to the reduced set of data points and therefore had to treat the finding with some caution.

[6] Truong & Weber (in prep.) used one female adult speaker and the same child speaker in the present study as speakers.

speakers. This time, four female adults and four children were used as speakers. Despite the increase in speakers, the results remained robust, meaning that the results found for the female adult speaker were indeed not due to a single speaker effect. This result also revealed that age of the speaker rather than gender caused the speaker effect for female participants. It should be noted that the two speakers of the present study differed not only in age (adult versus child) but also in gender (male adult versus female child). Based on the present study alone, we can thus not exclude the possibility that the gender of the speaker rather than age caused the speaker effect for female participants. Sexual dimorphism can be present in children as young as four years of age. Also typically, boys, at the age of seven or eight, have somewhat lower formant frequencies in comparison to girls (Vorperian & Kent, 2007). If gender rather than age of the speaker influences credibility ratings, then we should expect no difference in ratings between a female adult speaker and female child. It is therefore possible that participants identified the gender of our child speaker in the present study. This is, however, not what was found in Truong & Weber (in prep.), making it less likely that the present findings are driven by the gender of the speakers rather their age. When putting the exploratory findings of the present study together with the results of Truong & Weber (in prep.), it comes to light that both results rather lend support to the voice preference hypothesis, but it is not strictly about the influence of vocal pitch on credibility ratings, it is more about liking voices in general. If vocal pitch played a greater role in credibility judgements, then child speakers should have received lower credibility ratings than female adult speakers. The fact that female listeners rated female adult speakers as lower than children, makes it appear as if female listeners like children's and men's voices more than their own type of voices: Men's voices possibly because of their low pitch (e.g., Collins, 2000; Re et al., 2012) and children because of reasons which are yet to be investigated.

Taken together, the findings suggest that processing difficulties with child speech were not responsible for the lower credibility ratings in the present study, since the same child recordings resulted in higher judgements than female adult recordings in Truong & Weber (in prep.). Hence, the difference in credibility ratings between the adult and the child speaker in both studies probably emerged through listeners' voice preference. As the reduced set of data in the exploratory post-hoc analysis of the present study resulted in a weak effect, the interpretation has to be approached with caution and at this point is thus inevitably rather speculative. Nonetheless, the findings of the present study raise intriguing questions for future investigation. As the present work has solely comprised of a female child speaker, one further interesting avenue for future research would be to examine whether perceived credibility differs when employing a male child as the speaker. Yet another area of enquiry for future investigation is to employ younger children as speakers to further elucidate the relationship between speaker age and truth judgements. Furthermore, it remains open whether credibility ratings would differ if testing in a group of individuals who have extensive experience with child speakers (e.g., caregivers and preschool teachers) compared to a group who has no experience with children or rather limited experience to a particular group of children (e.g., parents). This would help us better understand whether the degree of exposure to children can potentially affect credibility ratings in general.

5   Conclusion

In sum, the present work investigated the relation between perceived credibility and speaker age by adult listeners. The overall analysis did not reveal the pattern we initially expected. Participants did not judge trivia statements as significantly less true when the statements were spoken by a child speaker than when the speaker was an adult. However, when taking participants' gender into account, different patterns of credibility judgement for both female and male participants were indicated in an exploratory analysis. For male listeners, there was no difference in the veracity of statements for both adult and child speakers. For female listeners, however only marginally significant, truth judgements were lower for the child speaker than for

the male adult speaker. We, therefore, interpret the results as an indirect relationship between truth judgement and speaker age, possibly mediated by voice preference.

References

Anderson, S., Downs, S. D., Faucette, K., Griffin, J., King, T., & Woolstenhulme, S. (2007). How accents affect perception of intelligence, physical attractiveness, and trustworthiness of middle-easter-, latin-american-, british- and standard-american-english-accented speakers. *Intuition: The BYU Undergraduate Journal in Psychology*, *3*(1), 5–11.

Bala, N., Ramakrishnan, K., Lindsay, R., & Lee, K. (2015). Judicial assessment of the credibility of child witnesses. *Alberta Law Review*, *42*, 995–1017.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using eigen and s4: R package.

Belin, P., Bestelmeyer, P., Latinus, M., & Watson, R. (2011). Understanding voice perception. *British Journal of Psychology*, *102*, 711–25.

Collins, S. A. (2000). Men's voices and women's choices. *Animal Behaviour*, *60*, 773–780.

Dewaele, J.-M., & McCloskey, J. (2015). Attitudes towards foreign accents among adult multilingual language users. *Journal of Multilingual and Multicultural Development*, *36*(3), 221–238.

Dion, K., Berscheid, E., & Walster, E. (1972). What is beautiful is good. *Journal of Personality and Social Psychology*, *24*(3), 285–290.

Dixon, J., Mahoney, B., & Cocks, R. (2002). Accents of guilt? effects of regional accent, race, and crime type on attributions of guilt. *Journal of Language and Social Psychology*, *21*, 162–168.

Edwards, J. (1999). Refining our understanding of language attitudes. *Journal of Language and Social Psychology*, *18*(1), 101–110.

Feinberg, D., Jones, B., Little, A., Burt, D., & Perrett, D. (2005). Manipulation of fundamental and formant frequencies influence the attractiveness of human male voices. *Animal Behaviour*, *69*, 561–568.

Ferguson, M. J., & Zayas, V. (2009). Automatic evaluation. *Current Directions in Psychological Science*, *18*(6), 362–366.

Fitch, W. (2000). The evolution of speech: A comparative review. *Trends in Cognitive Sciences*, *4*(7), 258–267.

Frumkin, L. A., & Stone, A. (2020). Not all eyewitnesses are equal: Accent status, race and age interact to influence evaluations of testimony. *Journal of Ethnicity in Criminal Justice*, *18*(2), 123–145.

Fuertes, J., Gottdiener, W., Martin, H., Gilbert, T., & Giles, H. (2012). A meta-analysis of the effects of speakers' accents on interpersonal evaluations. *European Journal of Social Psychology*, *42*, 120 – 133.

Fuertes, J., Potere, J., & Ramirez, K. (2002). Effects of speech accents on interpersonal evaluations: Implications for counseling practice and research. *Cultural diversity & ethnic minority psychology*, *8*, 346–56.

Giles, H. (1970). Evaluative reactions to accents. *Educational Review*, *22*(3), 211–227.

Gluszek, A., & Dovidio, J. F. (2010). The way they speak: A social psychological perspective on the stigma of nonnative accents in communication. *Personality and Social Psychology Review*, *14*(2), 214–237. PMID: 20220208.

Hughes, S. M., & Miller, N. E. (2015). What sounds beautiful looks beautiful stereotype: The matching of attractiveness of voices and faces. *Journal of Social and Personal Relationships*, *33*(7), 984–996.

Kent, R. (1976). Anatomical and neuromuscular maturation of the speech mechanism: Evidence from acoustic studies. *Journal of speech and hearing research*, *19*, 421–47.

Lee, S., Potamianos, A., & Narayanan, S. (1999). Acoustics of children's speech: Developmental changes of temporal and spectral parameters. *The Journal of the Acoustical Society of America*, *105*(3), 1455–1468.

Lenth, R. (2018). emmeans: Estimated marginal means, aka least-squares means. r package version 1.2.4.

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

Lindemann, S. (2003). Koreans, chinese or indians? attitudes and ideologies about non-native english speakers in the united states. *Journal of Sociolinguistics*, *7*, 348–364.

McAleer, P., Todorov, A., & Belin, P. (2014). How do you say 'hello'? personality impressions from brief novel voices. *PLoS One*, *9*, e90779.

McGlone, M. S., & Tofighbakhsh, J. (2000). Birds of a feather flock conjointly (?): Rhyme as reason in aphorisms. *Psychological Science*, *11*(5), 424–428.

Meo, A. D. (2012). How credibly is a non-native speaker? prosody and surroundings. In Busà Maria Grazia, S. A. (Ed.), *Methodological Perspectives on Second Language Prosody*, (pp. 3–9). Papers from ML2P 2012.

Munro, M. J., & Derwing, T. M. (1995). Processing time, accent, and comprehensibility in the perception of native and foreign-accented speech. *Language and Speech*, *38*, 289–306.

Munro, M. J., Derwing, T. M., & Satō, K. (2006). Salient accents, covert attitudes: Consciousness-raising for pre-service second language teachers.

Oppenheimer, D. M. (2008). The secret life of fluency. *Trends in Cognitive Sciences*, *12*(6), 237–241.

R Core Team (2013). R: A language and environment for statistical computing.

Re, D. E., O'Connor, J. J. M., Bennett, P. J., & Feinberg, D. R. (2012). Preferences for very low and very high voice pitch in humans. *PLoS One*, *7*(3), 1–8.

Reber, R., & Schwarz, N. (1999). Effects of perceptual fluency on judgments of truth. *Consciousness and Cognition*, *8*(3), 338 –342.

Reber, R., Schwarz, N., & Winkielman, P. (2004). Processing fluency and aesthetic pleasure: Is beauty in the perceiver's processing experience? *Personality and Social Psychology Review*, *8*(4), 364–382. PMID: 15582859.

Rezlescu, C., Penton, T., Walsh, V., Tsujimura, H., Scott, S. K., & Banissy, M. J. (2015). Dominant voices and attractive faces: The contribution of visual and auditory information to integrated person impressions. *Journal of Nonverbal Behavior*, *39*(4), 355–370.

Sell, A., Bryant, G. A., Cosmides, L., Tooby, J., Sznycer, D., von Rueden, C., Krauss, A., & Gurven, M. (2010). Adaptations in humans for assessing physical strength from the voice. *Proceedings of the Royal Society B*, *277*, 3509–3518.

Smith, B., Sugarman, M., & Long, S. (1983). Experimental manipulation of speaking rate for studying temporal variability in children's speech. *The Journal of the Acoustical Society of America*, *74*, 744–9.

Song, H., & Schwarz, N. (2009). If it's difficult to pronounce, it must be risky: Fluency, familiarity, and risk perception. *Psychological Science*, *20*(2), 135–138. PMID: 19170941.

Souza, A. L., & Markman, A. B. (2013). Foreign accent does not influence cognitive judgments. In *CogSci*.

Stocker, L. (2017). The impact of foreign accent on credibility: An analysis of cognitive statement ratings in a swiss context. *Journal of Psycholinguistic Research*, *46*, 617–628.

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

Tingley, B. M., & Allen, G. D. (1975). Development of speech timing control in children. *Child Development*, *46*(1), 186–194.

Truong, T. L., & Weber, A. (in prep.). The trust trap: Investigating trustworthiness of children and adults.

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. PMID: 27081101.

Unkelbach, C. (2006). The learned interpretation of cognitive fluency. *Psychological Science*, *17*(4), 339–345.

Vorperian, H., & Kent, R. (2007). Vowel acoustic space development in children: A synthesis of acoustic and anatomic data. *Journal of speech, language, and hearing research : JSLHR*, *50*, 1510–45.

Whittlesea, B. W., Jacoby, L. L., & Girard, K. (1990). Illusions of immediate memory: Evidence of an attributional basis for feelings of familiarity and perceptual quality. *Journal of Memory and Language*, *29*(6), 716 – 732.

Winter, B. (2019). *Statistics for Linguists: An Introduction Using R*. Routledge.

Xu, Y., Lee, A., Wu, W.-L., Liu, X., & Birkholz, P. (2013). Human vocal attractiveness as signaled by body size projection. *PLoS One*, *8*, e62397.

Zuckerman, M., & Miyake, K. (1993). The attractive voice: What makes it so? *Journal of Nonverbal Behavior*, *17*, 119–135.