Developing voice perception: an overview of current research and models in affect and identity recognition in children and adults

Gaby Mahrholz, Phil McAleer & Mandy Norrbo
School of Psychology, University of Glasgow
Email: g.mahrholz.1@research.gla.ac.uk
Email: philip.mcaleer@glasgow.ac.uk
Email: 2261281n@student.gla.ac.uk

University home page: http://www.psy.gla.ac.uk/staff/index.php?id=GM005
University home page: https://www.gla.ac.uk/schools/psychology/staff/philipmcaleer/

The human voice is a significant source of social information with an important role in conveying not only the identify of a speaker (Lavan, Burton, Scott, & McGettigan, 2019; Lavan, Scott, & McGettigan, 2016), but also their intentions, beliefs, and emotional states (Banse & Scherer, 1996; Mahrholz, Belin, & McAleer, 2018; McAleer, Todorov, & Belin, 2014; Pell & Skorup, 2008). Consider, for example, the ease in which we can identify a loved one, and their mood, from a mere "Hello" on the telephone. The underlying mechanisms involved were originally proposed in the *auditory face* model of voice processing by Belin, Fecteau and Bedard (2004), based on Bruce and Young's face model (1986). After an initial layer of low-level acoustical analysis, information is extracted from voices in three separate but interconnected pathways: speech analysis, affect analysis, recognition units. However, despite the prevalence of this model, much of the research that has been used to support the model has focussed on participants of late adolescences or early adulthood, and as such important questions as to how we develop these skills in our youth, and how these abilities change as we age, are as yet unanswered. In this brief review, we highlight key questions that must be considered if we are to obtain an understanding of how voice processing develops and matures from childhood to adulthood. In doing so, we will only focus on the nodes of affect analysis and voice recognition, allowing more specialist groups to focus on the development of speech analysis.

Affect Analysis

It is suggested that emotion recognition from auditory stimuli develops and improves gradually during childhood (Chronaki, Hadwin, Garner, Maurage, & Sonuga-Barke, 2015; Sauter, Panattoni, & Happe, 2013), reaching adult-like consensus by early- to mid-adolescence (Aguert, Laval, Lacroix, Gil, & Le Bigot, 2013; Brosgole & Weisman, 1995; Chronaki et al., 2015; Zupan, 2015). However, recent findings of a continued maturation of the *social brain network* well into late adolescence (Kilford, Garrett, & Blakemore, 2016) suggest that more fine-tuning of emotional maturation could occur between adolescence and adulthood (Morningstar, Ly, Feldman, & Dirks, 2018).

Whilst statistical mean averages suggest an overall level of ability comparable to adult level, emotion recognition between childhood and adulthood may actually vary by emotion type, intensity, and modality. For example, adults outperform adolescents on recognition of sadness, fear, and anger from voices (Morningstar et al., 2018), yet the accurate recognition of sadness, in both vocal and facial domains, is suggested to develop at a later stage than happiness and anger (Chronaki et al., 2015; - though see Jack, Garrod, & Schyns, 2014 for discussion on temporal distinction of emotion). On the other hand, Nelson and Russell (2011), and Quam and Swingley (2012), found toddlers and pre-schoolers could recognise emotion from facial or body language cues but not from voice-only cues, with pre-schoolers better at recognising sadness than anger, happiness, and fear (Nelson & Russell, 2011). Finally, Chronaki (2015) showed that emotion intensity may also play a role in successful vocal emotion recognition. Children aged 10-11 were less accurate in emotion recognition compared to adults when listening to 50% and 75% morphs of emotional states (between emotion and neutral). As such, it is still unclear as to when and how an accurate representation of different emotions develop in childhood.

However, one factor to potentially influence emotion recognition, and perhaps explaining competing results, is the complexity of the task and the ability of children to understand and correctly apply emotional categories. Typically, emotion research with young participants employs paradigms with a subset of two to four of the Ekman and Friesen (1971) canonical emotions: happiness, sadness, anger, surprise, fear, and disgust (Chronaki et al., 2015; Morningstar et al., 2018; Nelson & Russell, 2011; Quam & Swingley, 2012). Whilst little evidence exists as to whether or not children perform better when fewer emotion types are involved, it can be speculated that fewer categories would mean less cognitive load, leading to improved accuracy. Furthermore, emotion word knowledge may be an influencing factor to emotion recognition. One suggestion is that representations of emotional concepts develop from a mono-dimensional focus (between positive and negative) to a bi-dimensional focus (valence & arousal) between the ages of 6 and 25, and that increasing verbal knowledge mediates the development of emotion representation (Gendron, Lindquist, Barsalou, & Barrett, 2012; Nook, Sasse, Lambert, McLaughlin, &

Somerville, 2017). As such, varying results could be due to the age groups involved, the modality tested, and/or the complexity of the experimental paradigm.

A secondary element of affect analysis is perception of personality, reportedly linked to emotion through the overgeneralisation hypothesis (McArthur & Baron, 1983; Secord, 1958; Zebrowitz & Collins, 1997; Zebrowitz & Montepare, 2008). Though less researched than emotion, studies on vocal attractiveness have found that female adolescents and adults, but not children, perceive lower-pitched male voices as more attractive (Saxton, Caryl, & Roberts, 2006; Saxton, Debruine, Jones, Little, & Roberts, 2009, 2013). This suggests that perceived attractiveness changes during adolescence when mate-selection becomes important. However, younger but not older boys preferred higher-pitched female voices (Saxton et al., 2009, 2013), often rated as more attractive by adult men (e.g. Borkowska & Pawlowski, 2011), meaning sex and developmental stage may in fact interact across the lifespan.

Trustworthiness, a key trait in the *social voice space* model (McAleer et al., 2014), has been shown to elicit reliable judgements from adult listeners in both faces and voices, even after very brief exposure (for example, Mahrholz et al., 2018; McAleer et al., 2014; Oosterhof & Todorov, 2008; Willis & Todorov, 2006). However, how the percept of vocal trustworthiness develops and matures between childhood and adulthood is less clear. In the face research literature, an ability to judge how trustworthy emotionally neutral faces appear, develops early on in life (Caulfield, Ewing, Bank, & Rhodes, 2016; Cogsdill, Todorov, Spelke, & Banaji, 2014; Ewing, Caulfield, Read, & Rhodes, 2015). Given the similarities between the visual and auditory modality, we would expect vocal trustworthiness to emerge during childhood as well. Though at what age adult-like consensus is reached may depend on task difficulty and study paradigm, similarly to emotion research.

Overall, much work remains to establish how we develop and utilise concepts of emotions and personalities from voices. From face research we can propose that different emotions establish at different rates but with a lack of standardised approaches, it is unclear as to what this time-frame is and whether it is consistent with face research, or whether one modality benefits one percept over another. By taking a more systematic approach, using a consistent set of stimuli arrays, and treating age as the continuous variable that it is, we may start to draw a more fine-grained understanding of the development of perceiving emotion and personality from voices.

Voice Identity Recognition

To form a stable percept of a speaker's identity, a listener must determine whether differences between utterances are better explained by between-speaker (interspeaker) variability or within-speaker (intraspeaker) variability. Therefore, the process of perceiving identity in voices can be divided into two sub-processes, *telling people apar*t i.e. successful discrimination between two different people, and *telling people together* (Burton, 2013) i.e. recognition of when two utterances originate from the same person. Whereas acoustic differences between speakers have been studied for some time, research into the extent and nature of within-speaker variability has only recently started emerging (see Lavan, Burton, et al., 2019). As rightly argued by Lavan et al., (2019), to fully understand the perceptual processes involved in voice identity perception, it is essential to include within-speaker variability into future voice identity experiments and models.

An individual speaker's voice is subject to both naturally occurring variation as well as intentional changes made by the speaker. Natural variation may occur as a result of physiological and psychological changes as well as environmental stress. For example, there is evidence that changes in emotional state (Ozseven, Dugenci, Doruk, & Kahraman, 2018), hydration levels (Alves, Kruger, Pillay, van Lierde, & van der Linde, 2019), phase of the menstrual cycle (Banai, 2017), and prolonged use (Laukkanen & Kankare, 2006) may impact an individual's voice through changes in acoustic properties. Intentional variation is used to adapt to the social context (in a library versus in a loud concert), who we are speaking to (a child versus a close friend), and what emotion we are trying to convey (irony versus telling an exciting story). Even familiar speakers can become very difficult to identify if they intentionally alter their voices. In fact, recognition rates for familiar voices can be found to decrease from 97% in normal undisguised voices to 4% when utterances are spoken in falsetto (Wagner & Koster, 1999) highlighting substantial scope for within-speaker variability. Likewise, Lavan and colleagues (Lavan, Burston, & Garrido, 2019), using a sorting task paradigm, showed that introducing within-speaker variability in voice identity experiments reduces accuracy of recognition, especially for listeners who are unfamiliar with the voices: unfamiliar listeners categorise any perceived difference between two utterances as two different speakers rather than recognising that differences could be the result of within-speaker variability. As such, considering the potential variation within an individual voice, and that no two utterances from the same speaker ever sound the same (Van Lancker Sidtis & Kreiman, 2011), models and theories of voice identity perception that do not account for within-speaker variability are missing a large part of the picture.

Though little research currently exists that specifically focuses on how children recognise identity from voices, one area that has received focus is that of developmental disorders, such as Autism Spectrum Disorder (ASD), which appear to affect the ability to discriminate between others. These deficits are more evidenced in the face identity literature (cf. Weigelt, Koldewyn, & Kanwisher, 2012), however, there is a growing body of research indicating reduced voice identity perception abilities in children and adults diagnosed with ASD (Boucher, Lewis, & Collis, 1998; Schelinski, Riedel, & von Kriegstein, 2014; Schelinski, Roswandowitz, & von Kriegstein, 2017). Additionally, autistic traits, i.e. higher scores on the autism quotient (AQ), in the general population, have also been associated with reduced performance in voice identity perception tasks (Skuk, Palermo, Broemer, & Schweinberger, 2019). These findings are important as deficits in person identity processing can lead to social interaction difficulties, anxiety and avoidance of social situations (Yardley, McDermott, Pisarski, Duchaine, & Nakayama, 2008). Whilst the above psychosocial consequences were found in relation to deficits in face identity processing, it can be assumed that they also apply to situations in which voice identity becomes important, e.g. on the telephone.

There are several reasons why individuals diagnosed with ASD may display reduced performance in voice identity tasks. For example, there may be perceptual impairments in vocal pitch perception, and associated difficulties merging acoustical signals into a stable and coherent perception of a voice. Schelinski et al., (2017), for example, found evidence for such a deficit, as opposed to a deficit of linking voice to identity, in a sample of high-functioning ASD participants. Relatedly, irregular function in neural regions involved in voice identity perception have also been demonstrated in both people with ASD (Schelinski, Borowiak, & von Kriegstein, 2016) and people with developmental phonagnosia (Roswandowitz, Schelinski, & von Kriegstein, 2017); a rare disorder where patients display an inability to recognise familiar voices and impaired discrimination between voices. Contrary to neurodegenerative and lesion-induced forms of phonagnosia, developmental phonagnosia shows no impairment of cortical structures (Garrido et al., 2009). However, more studies are needed to understand the extent and nature of developmental differences in voice identity perception as our knowledge this far is hampered by the limited scope of the available research.

Overall, the field in general has largely neglected the role of within-speaker variability, focussing mainly on between-speaker variability for one reason or another. However, as a result of this knowledge gap, we currently know little as to how children, outwith those with a developmental disorder, cope with recognition of identifying people from their voice in the presence of other speakers. To better understand developmental difficulties and deficits in voice identity perception, we must first look to further elucidate the processes and steps involved in voice identity perception, incorporating studies with participants across the lifespan. From there, we will start to understand how we utilise strategies for voice identity perception, how these adapt from voice to voice, from listener to listener, and how these individual differences explain the discrepancies in voice recognition ability between neurotypicals and those with developmental disorders.

Competing Models of Perception

At the heart of the above research lies the overarching puzzle of how identities and emotions develop and are stored in the brain. Emotion research, has largely focussed on finite emotions (Ekman & Friesen, 1971), once thought to be culturally universal, and recognised through rigid, distinct, innate, neural pathways. Recent work now suggests that these are largely Westernised stereotypes (Chen & Jack, 2017; Jack et al., 2014) as evidenced via high cross-cultural variability in recognition ability. And whilst the Ekman emotions do give us understanding of cultural stereotypical norms, recent theoretical arguments maintain that each emotion, identity, and personality trait are not stored as one rigid structure ready for comparison when a decision is required.

Allowing for more flexibility in internal representation, voice identity, as well as emotion and personality from voices, has thus far been explained within the framework of a prototype-based model (Latinus & Belin, 2011). Such a model incorporates a fluid norm-based coding of voices around an average representation of an individual's voice in relation to a population voice space (Lavan, Knight, & McGettigan, 2019). When a new voice is encountered, the population voice space is used to make new/old (familiar/unfamiliar) judgements of the voice based on the similarity or dissimilarity of that stimulus to the internal norm. With more exposure to variable utterances from the new identity, incorporating a feedback system, the voice space that represents that identity may become more diverse, extended, or fluid, and as such more useful for managing previously unheard variability (Lavan, Knight, Hazan, & McGettigan, 2019). Further support for a flexible prototype model comes from the finding that individual voice spaces are of similar structure to population voice spaces, so that listeners can map each voice in relation to the population prototype (Lee, Keating, & Kreiman, 2019).

Barrett and colleagues (2019) offer an extensive review of several models varying in their understanding of how different traits are stored across neuronal firing, and how flexible the underlying prototypes are. For example, the norm-based coding model of voices lacks definition in the developmental process, underlining how we form a concept of certain percepts. Alternatively, a proposed prediction model (Barrett, 2017) suggests that there is no stored prototype within the brain, flexible or not, and instead percepts of emotion, personality, and identity, are constructed in each instance, when required, through statistical probability and predictions based on past experience, via an intersecting cascade of bottom-up and top-down neuronal firing. If this is indeed the case then, arguably, there is no innate sense of emotion or personality, as once previously thought, and children are merely hard-wired to be rapid statistical learners (Xu & Kushnir, 2013). Work by Denison and Xu (2010) would appear to support this idea, finding that infants as young as ten months old can show preferences based on estimated probabilities. Likewise, Mehr, Song and Spelke (2016) showed that 5-month-old infants will show a preference for a novel female singing a melody that they, the child, have become familiar to through their parents, proposing this as evidence for how we develop an understanding of our social network. In short, infants, through statistical probability, determine that the novel person singing the same song as their parents must be of the same group and therefore approachable.

Whether or not this prediction model is indeed how we develop concepts of emotions, identities and personalities, is a compelling suggestion but it remains to be tested. Degeneracy in brain networks (Edelman & Gally, 2001), that multiple neural pathways lead to the same percept, would all but rule out the idea of rigid stereotypes used for a continual game of pair matching. Yet, whilst the research does suggest that children can base judgements on a combination of probability and prediction, key questions remain, as highlighted above, as to how our internal representation models develop and change with time and ageing. These questions will undeniably keep cognitive and perception researchers occupied for the foreseeable future in both the behavioural and neural domains.

Conclusion

The *auditory face* model of voice processing (Belin et al., 2004) has been important in establishing a conceptual understanding of how we garner various strands of information from a speaker. However, it does not propose how we arrive at the conclusion that our perception matches the signal, or indeed how we first develop the ability to extract this information. Whilst the functioning behind these issues is currently debated between norm-based coding (Latinus & Belin, 2011) and prediction theory models (Barrett, 2017; Barrett et al., 2019), by theoretically combining the models we start to get a realistic sense of the process of extracting information from voices. Yet, as highlighted above, to fully harness this information, and to make use of it within the technological age, given the influx of attention towards voice-activated systems and Social Robotics (de Melo, Marsella, & Gratch, 2019; Mills, Bunnell, & Patel, 2014; Ponsot, Burred, Belin, & Aucouturier, 2018), we would do well to extend the field of voice research into both younger and older age ranges, focussing on such questions as proposed above, neglected in older ages as they are in younger ages. Only through doing so will we be able to arrive at a complete model of voice processing in the developing human.

Acknowledgements

Gaby Mahrholz is funded through a 1+3 Economic and Social Research Council (ESRC) award from the Scottish Graduate School of Social Science (SGSSS) [ES/J500136/1]. Mandy Norrbo was funded through the British Psychological Society Undergraduate Research Assistantship Scheme, Summer 2019.

References

Aguert, M., Laval, V., Lacroix, A., Gil, S., & Le Bigot, L. (2013). Inferring Emotions from Speech Prosody: Not So Easy at Age Five. *Plos One, 8*(12). doi:e8365710.1371/journal.pone.0083657

Alves, M., Kruger, E., Pillay, B., van Lierde, K., & van der Linde, J. (2019). The Effect of Hydration on Voice Quality in Adults: A Systematic Review. *Journal of Voice, 33*(1). doi:125.e1310.1016/j.jvoice.2017.10.001

Banai, I. P. (2017). Voice in different phases of menstrual cycle among naturally cycling women and users of hormonal contraceptives. *Plos One, 12*(8). doi:e018346210.1371/journal.pone.0183462

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

Barrett, L. F. (2017). The theory of constructed emotion: an active inference account of interoception and categorization. *Social Cognitive and Affective Neuroscience, 12*(1), 1-23. doi:10.1093/scan/nsw154

Barrett, L. F., Adolphs, R., Marsella, S., Martinez, A. M., & Pollak, S. D. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. *Psychological Science in the Public Interest, 20*(1), 1-68. doi:10.1177/1529100619832930

Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: neural correlates of voice perception. *Trends Cogn Sci, 8*(3), 129-135. doi:10.1016/j.tics.2004.01.008

Borkowska, B., & Pawlowski, B. (2011). Female voice frequency in the context of dominance and attractiveness perception. *Animal Behaviour, 82*(1), 55-59. doi:10.1016/j.anbehav.2011.03.024

Boucher, J., Lewis, V., & Collis, G. (1998). Familiar face and voice matching and recognition in children with autism. *Journal of Child Psychology and Psychiatry and Allied Disciplines, 39*(2), 171-181. doi:Doi 10.1017/S0021963097001820

Brosgole, L., & Weisman, J. (1995). Mood Recognition across the Ages. *International Journal of Neuroscience, 82*(3-4), 169-189. doi:Doi 10.3109/00207459508999800

Bruce, V., & Young, A. (1986). Understanding Face Recognition. *British Journal of Psychology, 77*, 305-327. doi:DOI 10.1111/j.2044-8295.1986.tb02199.x

Burton, A. M. (2013). Why has research in face recognition progressed so slowly? The importance of variability. *Quarterly Journal of Experimental Psychology, 66*(8), 1467-1485. doi:10.1080/17470218.2013.800125

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, 107*(3), 503-518. doi:10.1111/bjop.12156

Chen, C., & Jack, R. E. (2017). Discovering cultural differences (and similarities) in facial expressions of emotion. *Curr Opin Psychol, 17*, 61-66. doi:10.1016/j.copsyc.2017.06.010

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *Br J Dev Psychol, 33*(2), 218-236. doi:10.1111/bjdp.12075

Cogsdill, E. J., Todorov, A. T., Spelke, E. S., & Banaji, M. R. (2014). Inferring character from faces: a developmental study. *Psychol Sci, 25*(5), 1132-1139. doi:10.1177/0956797614523297

de Melo, C. M., Marsella, S., & Gratch, J. (2019). Human Cooperation When Acting Through Autonomous Machines. *Proceedings of the National Academy of Sciences of the United States of America, 116*(9), 3482-3487. doi:10.1073/pnas.1817656116

Denison, S., & Xu, F. (2010). Twelve- to 14-month-old infants can predict single-event probability with large set sizes. *Developmental Science, 13*(5), 798-803. doi:10.1111/j.1467-7687.2009.00943.x

Edelman, G. M., & Gally, J. A. (2001). Degeneracy and complexity in biological systems. *Proceedings of the National Academy of Sciences of the United States of America, 98*(24), 13763-13768. doi:DOI 10.1073/pnas.231499798

Ekman, P., & Friesen, W. V. (1971). Constants across Cultures in Face and Emotion. *Journal of Personality and Social Psychology, 17*(2), 124-&. doi:DOI 10.1037/h0030377

Ewing, L., Caulfield, F., Read, A., & Rhodes, G. (2015). Perceived trustworthiness of faces drives trust behaviour in children. *Developmental Science, 18*(2), 327-334. doi:10.1111/desc.12218

Garrido, L., Eisner, F., McGettigan, C., Stewart, L., Sauter, D., Hanley, J. R., . . . Duchaine, B. (2009). Developmental phonagnosia: A selective deficit of vocal identity recognition. *Neuropsychologia, 47*(1), 123-131. doi:10.1016/j.neuropsychologia.2008.08.003

Gendron, M., Lindquist, K. A., Barsalou, L., & Barrett, L. F. (2012). Emotion Words Shape Emotion Percepts. *Emotion, 12*(2), 314-325. doi:10.1037/a0026007

Jack, R. E., Garrod, O. G. B., & Schyns, P. G. (2014). Dynamic Facial Expressions of Emotion Transmit an Evolving Hierarchy of Signals over Time. *Current Biology, 24*(2), 187-192. doi:10.1016/j.cub.2013.11.064

Kilford, E. J., Garrett, E., & Blakemore, S.-J. (2016). The development of social cognition in adolescence: An integrated perspective. *Neuroscience & Biobehavioral Reviews, 70*, 106-120. doi:https://doi.org/10.1016/j.neubiorev.2016.08.016

Latinus, M., & Belin, P. (2011). Anti-voice adaptation suggests prototype-based coding of voice identity. *Frontiers in Psychology, 2*. doi:10.3389/fpsyg.2011.00175

Laukkanen, A. M., & Kankare, E. (2006). Vocal loading-related changes in male teachers' voices investigated before and after a working day. *Folia Phoniatrica Et Logopaedica, 58*(4), 229-239. doi:10.1159/000093180

Lavan, N., Burston, L. F. K., & Garrido, L. (2019). How many voices did you hear? Natural variability disrupts identity perception from unfamiliar voices. *British Journal of Psychology, 110*(3), 576-593. doi:10.1111/bjop.12348

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

Lavan, N., Knight, S., Hazan, V., & McGettigan, C. (2019). The effects of high variability training on voice identity learning. *Cognition, 193*, 104026. doi:10.1016/j.cognition.2019.104026

Lavan, N., Knight, S., & McGettigan, C. (2019). Listeners form average-based representations of individual voice identities. *Nature Communications, 10*. doi:240410.1038/s41467-019-10295-w

Lavan, N., Scott, S. K., & McGettigan, C. (2016). Impaired Generalization of Speaker Identity in the Perception of Familiar and Unfamiliar Voices. *Journal of Experimental Psychology-General, 145*(12), 1604-1614. doi:10.1037/xge0000223

Lee, Y., Keating, P., & Kreiman, J. (2019). Acoustic voice variation within and between speakers. *Journal of the Acoustical Society of America, 146*(3), 1568-1579. doi:10.1121/1.5125134

Mahrholz, G., Belin, P., & McAleer, P. (2018). Judgements of a speaker's personality are correlated across differing content and stimulus type. *Plos One, 13*(10). doi:e020499110.1371/journal.pone.0204991

McAleer, P., Todorov, A., & Belin, P. (2014). How Do You Say 'Hello'? Personality Impressions from Brief Novel Voices. *Plos One, 9*(3). doi:e9077910.1371/journal.pone.0090779

McArthur, L. Z., & Baron, R. M. (1983). Toward an Ecological Theory of Social-Perception. *Psychological Review, 90*(3), 215-238. doi:Doi 10.1037/0033-295x.90.3.215

Mehr, S. A., Song, L. A., & Spelke, E. S. (2016). For 5-Month-Old Infants, Melodies Are Social. *Psychological Science, 27*(4), 486-501. doi:10.1177/0956797615626691

Mills, T., Bunnell, H. T., & Patel, R. (2014). Towards Personalized Speech Synthesis for Augmentative and Alternative Communication. *Augmentative and Alternative Communication, 30*(3), 226-236. doi:10.3109/07434618.2014.924026

Morningstar, M., Ly, V. Y., Feldman, L., & Dirks, M. A. (2018). Mid-Adolescents' and Adults' Recognition of Vocal Cues of Emotion and Social Intent: Differences by Expression and

Speaker Age. *Journal of Nonverbal Behavior, 42*(2), 237-251. doi:10.1007/s10919-018-0274-7

Nelson, N. L., & Russell, J. A. (2011). Preschoolers' use of dynamic facial, bodily, and vocal cues to emotion. *Journal of Experimental Child Psychology, 110*(1), 52-61. doi:10.1016/j.jecp.2011.03.014

Nook, E. C., Sasse, S. F., Lambert, H. K., McLaughlin, K. A., & Somerville, L. H. (2017). Increasing verbal knowledge mediates development of multidimensional emotion representations. *Nature Human Behaviour, 1*(12), 881-889. doi:10.1038/s41562-017-0238-7

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

Ozseven, T., Dugenci, M., Doruk, A., & Kahraman, H. I. (2018). Voice Traces of Anxiety: Acoustic Parameters Affected by Anxiety Disorder. *Archives of Acoustics, 43*(4), 625-636. doi:10.24425/aoa.2018.125156

Pell, M. D., & Skorup, V. (2008). Implicit processing of emotional prosody in a foreign versus native language. *Speech Communication, 50*(6), 519-530. doi:10.1016/j.specom.2008.03.006

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. doi:10.1073/pnas.1716090115

Quam, C., & Swingley, D. (2012). Development in Children's Interpretation of Pitch Cues to Emotions. *Child Development, 83*(1), 236-250. doi:10.1111/j.1467-8624.2011.01700.x

Roswandowitz, C., Schelinski, S., & von Kriegstein, K. (2017). Developmental phonagnosia: Linking neural mechanisms with the behavioural phenotype. *Neuroimage, 155*, 97-112. doi:10.1016/j.neuroimage.2017.02.064

Sauter, D. A., Panattoni, C., & Happe, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology, 31*(1), 97-113. doi:10.1111/j.2044-835X.2012.02081.x

Saxton, T. K., Caryl, P. G., & Roberts, S. C. (2006). Vocal and facial attractiveness judgments of children, adolescents and adults: The ontogeny of mate choice. *Ethology, 112*(12), 1179-1185. doi:10.1111/j.1439-0310.2006.01278.x

Saxton, T. K., Debruine, L. M., Jones, B. C., Little, A. C., & Roberts, S. C. (2009). Face and voice attractiveness judgments change during adolescence. *Evolution and Human Behavior, 30*(6), 398-408. doi:10.1016/j.evolhumbehav.2009.06.004

Saxton, T. K., DeBruine, L. M., Jones, B. C., Little, A. C., & Roberts, S. C. (2013). Voice pitch preferences of adolescents: Do changes across time indicate a shift towards potentially adaptive adult-like preferences? *Personality and Individual Differences, 55*(2), 90-94. doi:10.1016/j.paid.2013.02.009

Schelinski, S., Borowiak, K., & von Kriegstein, K. (2016). Temporal voice areas exist in autism spectrum disorder but are dysfunctional for voice identity recognition. *Social Cognitive and Affective Neuroscience, 11*(11), 1812-1822. doi:10.1093/scan/nsw089

Schelinski, S., Riedel, P., & von Kriegstein, K. (2014). Visual abilities are important for auditory-only speech recognition: Evidence from autism spectrum disorder. *Neuropsychologia, 65*, 1-11. doi:10.1016/j.neuropsychologia.2014.09.031

Schelinski, S., Roswandowitz, C., & von Kriegstein, K. (2017). Voice identity processing in autism spectrum disorder. *Autism Research, 10*(1), 155-168. doi:10.1002/aur.1639

Secord, P. F. (1958). The Social Stereotype and the Concept of Implicit Personality Theory. *American Psychologist, 13*(7), 329-329.

Skuk, V. G., Palermo, R., Broemer, L., & Schweinberger, S. R. (2019). Autistic Traits are Linked to Individual Differences in Familiar Voice Identification. *Journal of Autism and Developmental Disorders, 49*(7), 2747-2767. doi:10.1007/s10803-017-3039-y

Van Lancker Sidtis, D., & Kreiman, J. (2011). In the Beginning Was the Familiar Voice: Personally Familiar Voices in the Evolutionary and Contemporary Biology of Communication. *Integrative psychological & behavioral science, 46*, 146-159. doi:10.1007/s12124-011-9177-4

Wagner, I., & Koster, O. (1999). *Perceptual recognition of familiar voices using falsetto as a type of disguise.* Paper presented at the 14th International Congress of Phonetic Sciences, San Francisco.

Weigelt, S., Koldewyn, K., & Kanwisher, N. (2012). Face identity recognition in autism spectrum disorders: A review of behavioral studies. *Neuroscience and Biobehavioral Reviews, 36*(3), 1060-1084. doi:10.1016/j.neubiorev.2011.12.008

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. doi:DOI 10.1111/j.1467-9280.2006.01750.x

Xu, F., & Kushnir, T. (2013). Infants Are Rational Constructivist Learners. *Current Directions in Psychological Science, 22*(1), 28-32. doi:10.1177/0963721412469396

Yardley, L., McDermott, L., Pisarski, S., Duchaine, B., & Nakayama, K. (2008). Psychosocial consequences of developmental prosopagnosia: A problem of recognition. *Journal of Psychosomatic Research, 65*(5), 445-451. doi:10.1016/j.jpsychores.2008.03.013

Zebrowitz, L. A., & Collins, M. A. (1997). Accurate social perception at zero acquaintance: the affordances of a Gibsonian approach. *Pers Soc Psychol Rev, 1*(3), 204-223. doi:10.1207/s15327957pspr0103_2

Zebrowitz, L. A., & Montepare, J. M. (2008). Social Psychological Face Perception: Why Appearance Matters. *Soc Personal Psychol Compass, 2*(3), 1497. doi:10.1111/j.1751-9004.2008.00109.x

Zupan, B. (2015). Recognition of high and low intensity facial and vocal expressions of emotion by children and adults. *Journal of Social Sciences and Humanities, 1*, 332-344.