



**HAL**  
open science

# Development of Voice Perception in the Human Brain

Marie-Hélène Grosbras, Pascal Belin

► **To cite this version:**

Marie-Hélène Grosbras, Pascal Belin. Development of Voice Perception in the Human Brain. Jean Decety. The Social Brain: A Developmental Perspective, 2020. hal-03084465

**HAL Id: hal-03084465**

**<https://hal.science/hal-03084465v1>**

Submitted on 21 Dec 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# I Early Social Perception and Cognition

\_\_\_-1  
\_\_\_0  
\_\_\_+1

-1\_\_  
0\_\_  
+1\_\_

# 1 Development of Voice Perception in the Human Brain

Marie-Hélène Grosbras and Pascal Belin

## Overview

Routinely and effortlessly we extract a wealth of socially relevant information from voices. This goes far beyond speech comprehension and allows us to recognize a variety of attributes from the speaker such as identity or affect. Here we review evidence showing that this expertise starts developing very early on in infancy and continues to refine throughout childhood and part of adolescence. We also examine the maturation of dedicated voice processing mechanisms in the brain. We highlight similarities with the development of face processing and discuss implications and future directions for research.

## Introduction

Identifying relevant features from voices is paramount not only for developing linguistic competences, but also for building sophisticated social interactions throughout the life span. Indeed, humans exhibit impressive skills for extracting subtle information from a voice, including identity, age, gender, and minute emotional or intentional expressions. An extensive corpus of studies has described the perceptual and neurophysiological processes specific to voices as compared to other sounds. This involves voice-sensitive brain regions, which are more active when listening to conspecific vocalizations than when listening to environmental sounds, centered primarily around the temporal voice areas (TVA) along the superior temporal cortex (see figure 1.1B), and including also prefrontal, limbic, and subcortical regions. These processes are probably phylogenetically ancient as homologous processes are present in other species. Their ontogeny is less known.

\_\_\_-1  
\_\_\_0  
\_\_\_+1

Yet the ability to perceive, categorize, and appraise complex stimuli such as voices in a vast range of contexts necessitates advanced expertise, which must be built as an individual grows up. The consolidation of this expertise is facilitated by the bias that even young infants show to attend to social cues, including faces or voices, more than other stimuli in their environment (Valenza, Simion, Cassia, & Umiltà, 1996; Vouloumanos, Hauser, Werker, & Martin, 2010). Being able to process vocal stimuli in a sophisticated way is obviously important for language acquisition. However, in addition—and often overlooked in the developmental literature—this ability is also fundamental for learning to interact with others. In this regard, voice has sometimes been compared to an “auditory face” (Belin, Fecteau, & Bedard, 2004), conveying cues about an individual traits and states of mind. Indeed, both voices and faces carry important social information; they are present early in the infant’s environment, and their perception also is modulated by context. One might thus expect a similar developmental trajectory for the perception of both channels of social interaction.

Extensive research has shown that face perception skills (including identity, emotion, and intention recognition) are present very early on in infancy (e.g., Walker-Andrews, 1997) and continue to improve during adolescence (Scherf, Behrmann, & Dahl, 2012). Likewise, brain activity in the fusiform cortex sensitive to face stimuli, which is present in young infants (Halit, Csibra, Volein, & Johnson, 2004), continues to mature during adolescence (Scherf et al., 2012). In comparison, the development of voice processing has been less investigated, and these studies have focused principally on infancy and early childhood. Yet, as with faces, expertise for voices continues to mature throughout childhood and adolescence, alongside improvement and complexification of verbal but also nonverbal communication. Here we present an overview of the main findings, at the behavioral and neural levels, concerning voice processing in infants, children, and adolescents.

## Development of Voice Perception Skills

### Infancy and Toddlerhood

Using measures of heart rate variability, several studies have provided evidence that near-term fetuses in utero are sensitive to familiar voices (Lee & Kisilevsky, 2014). At birth, infants prefer to orient toward their mother’s

voice (DeCasper & Fifer, 1980). This sensitivity to voice identity appears earlier than recognition of speech-related features and may facilitate language acquisition. Indeed, it is only from two months onward that infants can discriminate between different phonemes (Friederici, 2005), and their recognition improves if the phonemes are spoken by their mother than an unfamiliar speaker (Barker & Newman, 2004).

Early on, infants also can extract categorical identity attributes from voices. After using habituation paradigms, Miller (1983) reported that six-month-old infants, but not two-month-old infants, could distinguish male from female voices. Around the same period, infants show responses that appear different for adults' approving or prohibitive vocalizations, even if they are spoken in another language (Fernald, 1993); infants can discriminate happiness, anger, and sadness from the tone of voice (Flom & Bahrick, 2007). In the same study five-month-olds were found to achieve chance level for recognizing facial expressions. In fact, several studies have concluded that during the first year of life voice is the primary channel for recognizing emotion, even though infants demonstrate some knowledge of the matching between facial expressions and the corresponding affective vocal expressions (Walker-Andrews & Lennon, 1991). Accordingly, infants adapt their behavior primarily in response to the affect expressed by their mother's voice rather than face. For instance, Mumme, Fernald, and Herrera (1996) showed that twelve-month-olds interacted more with a novel toy if their mother orally expressed a positive rather than negative emotion. By contrast, facial expressions did not influence the child's behavior.

The earliest development of voice recognition skills compared to face recognition could be explained by several factors. First, during fetal development the auditory system is stimulated earlier than the visual system, giving a head start to vocal perception skills. Second, in the context of an immature orienting system, auditory processing becomes more relevant than visual processing in many contexts, when the interacting person is not directly in the field of view. Lastly, the evolutionary importance of expression of emotion from voice likely contributes to why voice processing or at least affective voice processing develops earlier than face processing in ontogeny.

### Childhood and Adolescence

Although babies and toddlers can extract identity, gender, and affective information from voice above chance level, they are still not as proficient as adults. Considerable improvement takes place during childhood and adolescence. Research on this topic is somewhat disjoint from that on infant development because the methods used to study preverbal individuals differ from those used in children, which are more similar to investigations in adults. Although voice is less researched than face perception, developmental studies have indicated that, similar to faces, the development of voice perception is different from the development of perception of other, nonsocial stimuli.

Also, children perform well at tasks directly testing voice recognition. In an early study by Bartholomeus (1973), four-year-old children were at 57 percent correct at recognizing (free naming) their kindergarten classmates from hearing their voice (compared with 68 percent for the teachers). They performed far better with face recognition (97 percent for children, 100 percent for adults). Not much was done in this research domain in the following years, when efforts focused on the nature and development of children's linguistic representations. The implicit conclusion was that little change took place after age four with regards to processing of paralinguistic information from voices.

Yet studies using more controlled designs and populations have demonstrated that identity recognition based on voices continues to improve during school years. Creel and Jimenez (2012) observed that children aged three to six years were better than chance when discriminating between speakers, even when the speakers were close in terms of their voice acoustic parameters, but the children were still far worse than adults (60 percent versus 90 percent correct responses). Mann, Diamond, and Carey (1979) showed that the recognition of recently learned voices improved from age six to ten, followed by a dip in performance during early adolescence before attainment of adult capacity by age fourteen. This was confirmed with a similar task using monosyllabic words instead of sentences. Levi and Schwartz (2013) showed that children (aged seven to nine years) performed worse than early adolescents (aged ten to twelve years), who were themselves not as good as adults.

The emerging picture is thus that preschoolers but also school-aged children are less proficient than adults at mapping fine-grained voice features

-1\_\_\_  
0\_\_\_  
+1\_\_\_

to individuals. Nevertheless, for more salient distinctions—such as male versus female speakers, or child versus adult voices—adult-level performance is achieved much earlier (Creel & Jimenez, 2012). Interestingly, no own-age bias was reported, similar to what has been described for faces in this age range (Macchi Cassia, Pisacane, & Gava, 2012).

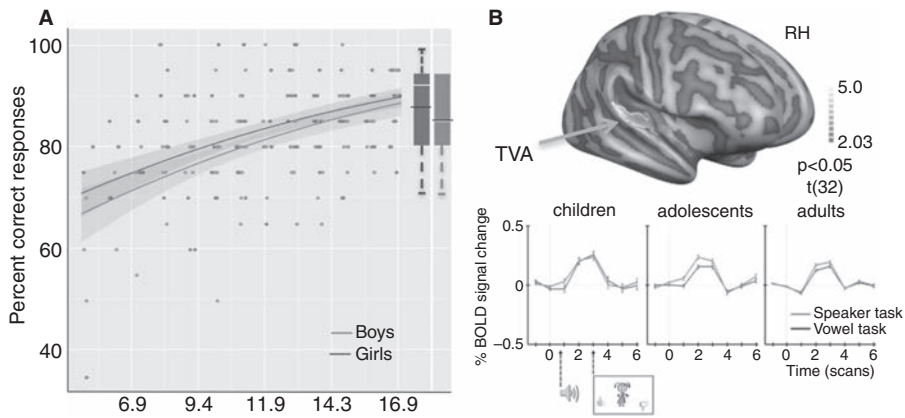
Affective state recognition from voice also improves considerably during childhood. The recognition of basic emotions from the tone of voice in sentences improves between five and ten years of age (Boucher, Lewis, & Collis, 2000; Cohen, Prather, Town, & Hynd, 1990; Friend, 2000; Sauter, Panatoni, & Happé, 2013; Van Lancker, Cornelius, & Kreiman, 1989). This is also the case when using meaningless sentences, single words (Matsumoto & Kishimoto, 1983), or nonverbal interjections or vocal bursts (Allgood & Heaton, 2015; Grosbras, Ross, & Belin, 2018; Sauter et al., 2013) as stimuli. This development is delayed compared with recognition of affective intent from music and from development of linguistic comprehension. This suggests a maturational trajectory distinct from language and from general emotion comprehension development. Rather, it seems related to a specific tuning of vocal cues decoding with age. It continues also during adolescence, until about fifteen years of age, with parallel trajectories for boys and girls (figure 1.1A) (Grosbras et al., 2018; Chronaki, Hadwin, Garner, Mauraage, & Sonuga-Barke, 2015; Morningstar, Ly, Feldman, & Dirks, 2018).

This developmental time line is comparable to what has been described from the literature on face perception (Herba & Phillips, 2004; Wade, Lawrence, Mandy, & Skuse, 2006). Also, studies directly comparing the two modalities in children have suggested—contrary to what has been observed in infants—an advantage of the face channel for recognizing identity (Bartholomeus, 1973) or emotion (Gil, Aguert, Le Bigot, Lacroix, & Laval, 2014; Nowicki & Mitchell, 1998; Zupan, 2015), similar to that observed in adults. Furthermore, if lexical (Friend, 2000) or contextual information including facial expression (Gil et al., 2014) is present, children aged three to nine years rely more on this information than on prosody to decipher the affective state of people.

In summary, the ability to extract information about the speaker's identity or emotion continues to mature until adolescence, after a significant amount of exposure to both familiar and novel voices. This resembles what has been described for the ability to extract the same information from faces. This late refinement of social perception is in line with social

\_\_\_-1  
\_\_\_0  
\_\_\_+1





**Figure 1.1**

A. Percentage correct in a four-alternative forced choice for emotion from nonlinguistic utterances from ages four to seventeen. The lines represent the predicted performances for males and females separately. Shaded areas represent confidence intervals. Boxplots represent median and 75 percentile in adults' data. (Reprinted with permission from Grosbras et al., 2018.) B. Functional magnetic resonance imaging response in the right TVA when children, adolescents, or adults judged the speaker's identity (light grey) or categorized the vowel pronounced from short utterances (dark grey). Note the highest signal for the speaker task for the adolescents and adults but not the children. TVA: temporal voice area. (Reprinted with permission from Bonte et al., 2016.)

challenges during the transition into adulthood: as social relationships become more complex, individuals need to be able to detect and categorize subtle social cues more accurately.

Life history theories posit that the adaptive value of adolescence as a specific stage in the human life span is to allow for better social skills building and that this is related to increased brain size and protracted development of the brain networks involved in social perception (Bogin, 1994; Joffe, 1997). This account raises the question of how brain development may support development in vocal decoding abilities. Indeed, regions of the temporal lobes, including the TVA, are among the latest to reach an adult level of structural maturity (Mills, Lalonde, Clasen, Giedd, & Blake-more, 2014). In the next section, we examine evidence of the maturation of functional brain activity related to voice perception throughout infancy, childhood, and adolescence.

## Development of Voice Perception in the Brain

### Infancy and Toddlerhood

Brain responses to vocal stimuli have been studied in infants and toddlers by use of electroencephalography (EEG), functional near-infrared spectroscopy (fNIRS), and more recently functional magnetic resonance imaging (fMRI). Although many studies have aimed at finding neural correlates of speech, a few have focused on voice perception *per se*, thereby using nonlinguistic stimuli or at least stimuli without any semantic content.

Beauchemin and colleagues (2011) measured EEG responses in newborns evoked by either the mother's voice or another familiar voice in contrast to unfamiliar voices, all uttering the vowel /a/ (mismatch-negativity paradigm). Differences in the response to the mother and stranger voices emerged 100 milliseconds after stimulus presentation, peaking at 200 milliseconds, with the higher signal and stronger left hemisphere dominance appearing for the mother's voice. Using a similar technique, Purhonen and colleagues (2004) observed that four-month-old infants displayed an increased negativity at 300 milliseconds after they heard their mother's as compared to a stranger's voice. Both fNIRS studies in neonates (Saito et al., 2007) and fMRI studies in two- to three-month-old infants (Dehaene-Lambertz et al., 2010) also showed changes in metabolic signal in the frontal cortex for the mothers compared with strangers' voices. Although it is difficult to say to what extent these observations are related to perceptual, attentional, or emotional effects, they do demonstrate that from the very first weeks of life, at an early stage of cortical auditory processing the timing, the intensity and topology of brain activity reflect some information about the speaker's identity, sufficient to differentiate the mother's voice from other voices.

Other brain imaging studies have shown sensitivity to prosody or emotional content of voice as well. For instance, Grossmann, Striano, and Friederici (2005) observed in seven-month-old infants a slow-wave event-related brain potential (ERP) over the temporal cortex from 500 milliseconds to 1,000 milliseconds after the onset of a word uttered with a happy or angry but not a neutral prosody. This was confirmed by a fNIRS study showing modulation of hemodynamic response in the right temporal and right prefrontal cortex when seven-month-olds listened to sentences spoken happily or angrily as compared to neutral prosody (Grossmann, Oberecker,

\_\_\_-1  
\_\_\_0  
\_\_\_+1

Koch, & Friederici, 2010). This suggests enhanced sensory processing for the affectively loaded auditory stimuli.

Using the mismatched negativity technique, which measures ERP signal in response to a “deviant” stimulus in a stream of stimuli from another category, Cheng and colleagues (2012) showed that even neonates display a response recorded from frontal-central electrodes sensitive to nonlinguistic happy and fearful prosody as compared to neutral sounds. In a follow-up study Zhang and colleagues (2017) showed increased hemodynamic response over the right middle and superior temporal cortex when neonates listened passively to emotional prosody (fearful angry or happy prosody in pseudo sentences—that is, sentences devoid of semantic meaning). Blasi and colleagues (2011), who used fMRI in seven-month-old infants, did not observe any significant difference between happy and neutral nonspeech vocalizations. Sad vocalizations, by contrast, elicited specific activation, but over the insula and orbitofrontal cortex and not over the superior temporal cortex.

Thus, at birth and in the first months, the brain is certainly sensitive to the affective content of a voice in addition to its familiarity, but it remains unclear exactly which emotions elicit the most differential signals. At any rate, this fits with the behavioral observations described in the previous section.

Yet although these studies show that speakers’ attributes such as identity or affective states are processed differently in the infant brain very early on, they do not show whether there is something specific about voice per se. The same responses could be observed when contrasting two musical instruments, for instance, or other auditory objects with different familiarity or symbolic value. Also, depending on the study, the brain responses are observed over the temporal or frontal cortex.

To address this question, Cheng and colleagues (2012) contrasted happy and angry vocalizations to acoustically matched synthetic sounds, and they observed a mismatch negativity at the electrodes compatible with the location of the TVA that was specific to vocal sounds, thus demonstrating higher temporal cortex activity for vocal as compared to nonvocal synthetic stimuli in neonates. Yet in an fNIRS study contrasting nonverbal vocalizations to natural environmental sounds (taken from the voice localizer commonly used in adults and including animal vocalizations), Grossmann and colleagues (2010) observed voice-specific activity in seven-month-olds but not in four-month-olds, albeit in a location more posterior than what would be expected from direct comparison with data from adults (Belin & Grosbras,

-1\_\_\_  
0\_\_\_  
+1\_\_\_

2010). Using fMRI and contrasting familiar nonlinguistic vocalizations and environmental sounds Blasi and colleagues (2011) reported activity in several foci along the superior temporal sulcus in three- to seven-month-old infants. These different results might be due to differences in methodology, but they also may reflect the fact that cortical voice sensitivity in early life might be less robust and less specific. In line with this hypothesis, Lloyd-Fox and colleagues (2012) reported voice-selective hemodynamics in the same location as well as in a slightly more anterior channel, with a linear increase between three and seven months of age.

Furthermore a fNIRS study found that four-month-old infants showed similarly strong responses in the temporal cortex on hearing monkey calls as they did when hearing human speech; this indicates that at this young age infants are still sensitive to vocalizations from both species (Minagawa-Kawai et al., 2011). This could also possibly explain why Grossman and colleagues (2010) did not observe any voice selective activity in four-month-olds, because their control stimuli included animal vocalizations. However, given the small number of brain imaging studies in infants, when exactly voice-sensitive or voice-selective responses appear in the cortex remains to be confirmed. What is certain is that specialized processes exist in the infant brain before the end of the first year that are dedicated to processing voice stimuli.

### Childhood and Adolescence

The research on voice brain processing in children and adolescents is even sparser. Using ERP, Rogier and colleagues (2010) observed that four- to five-year-old children displayed frontal evoked activity over frontotemporal channels between 200 and 300 milliseconds after stimulus onset specifically when they listened to voices, but not other auditory stimuli. This response is similar to the “frontal-temporal positivity to voice” described in adults (Charest et al., 2009). This result was replicated in four- to twelve-year-old children as well as in a group of children with cochlear implants, albeit in this case with a slightly later latency (Bakhos, Galvin, Roux, Lescanne, & Bruneau, 2018).

Using fMRI, Abrams and colleagues (2016) described voice-specific activity (while contrasting voices to environmental sounds) along the superior temporal sulcus in children aged seven to twelve years. Interestingly the anterior foci were active only when children listened to their mother’s

\_\_\_-1  
\_\_\_0  
\_\_\_+1

voice and not to unfamiliar voices, raising the possibility that at this age the activity is driven by the social or affective relevance of the vocal source (see also Yamasaki et al., 2013), or that the anterior temporal lobe stores memory for familiar voices. The study made no direct comparison with adults, however.

Using fMRI in children (aged eight to nine years), adolescents (aged fourteen to fifteen years), and young adults, Bonte et al. (2013) looked at the developmental trajectories of both the function and morphology of the superior temporal cortex region. They observed voice-selective activity in the three age-groups along the superior temporal sulcus and gyrus, corresponding to the TVAs. As in infants and adults, the activity was higher in the right hemisphere. Also this voice response in the superior temporal cortex changed from being less selective (lower  $t$  values) and more spatially diffuse (larger numbers of voxels) in children toward being highly selective and focal in adults. In a match-to-sample task on vowels uttered by children or adults, focusing either on identifying the vowel spoken or the speaker, the same authors observed higher activity for the speaker task compared to the vowel (speech) task in the left TVA in children adolescents and adults. This effect increased with age (figure 1.1A) (Bonte, Ley, Scharke, & Formisano, 2016).

Thus, the specialization of the temporal cortex for vocal stimuli occurs independent of the development of speech processing. Moreover, compared to adults and adolescents, children recruited additional regions in the cingulate and frontal cortex. Although they should be tested with connectivity analyses, these findings indicate that the functional network for processing voices becomes more focal with age.

In summary, the few studies investigating brain correlates of the paralinguistic aspects of voice processing show that they continue to mature in childhood and adolescence. Similarly to what has been described for other cognitive or perceptual processes, notably face perception, the voice cortical network might develop from a more diffuse, less specialized configuration to a more focal, more specialized one.

### Implications and Conclusions

The early presence of specialized processes for voice perception and their protracted refinement to reach adult expertise have implications for the maturation of how an individual interacts with her environment at

different moments of life. First, voice processing is one important channel for nonverbal communication, which has different constraints at different life stages. Preverbal infants are entirely dependent on caregivers and thus on social interactions. From birth, it is thus vital to learn quickly how to decode, adapt to, and send social signals. It makes sense that dedicated mechanisms exist in the brain very early on in development. Also, the intrinsic incentive salience and behavioral relevance of stimuli like faces and voices probably help to promote experience-dependent plasticity (Poremba, Bigelow, & Rossi, 2013) in specialized neural substrates, thereby increasing the specificity of response to these stimuli.

Furthermore, in view of the complexity of these signals and the subtlety of the information to be extracted, it is also not surprising that the refinement of social perception—and in particular voice perception—takes over a decade to complete. This notion fits with life history theories of evolution that propose that the emergence of adolescence as a distinct stage in ontogeny, between childhood and adulthood, serves the purpose of developing sophisticated social skills and thus promotes group cohesion (Bogin, 1994). As such, adolescence would have conferred significant reproductive advantages on our evolutionary ancestors, in part by allowing individuals to learn and master adult relationship skills before mating and reproducing. From this evolutionary perspective, the development of voice perception in adolescence could be related to the role of voice in courtship in many species.

The adaptive values of building advanced expertise in decoding social signals is exemplified by the correlation between performance in recognition tests of identity or emotion and social competencies, including popularity, engagement in conflict, conflict resolution, communication, and social interaction in preschool-aged children (Nowicki & Mitchell, 1998) and in children and adolescents (Abrams et al., 2016; Skuk, Palermo, Broemer, & Schweinberger, 2019), as well as academic achievement. In studies that also tested facial expression recognition, social competence was associated more strongly with vocal than facial emotion recognition (Collins & Nowicki, 2001). This might relate to the fact that, as in adults, in children and adolescents performance in identity and emotion recognition becomes more challenging with voices than faces (Stevenage et al., 2013). This contrasts with early life when independent movement is limited and vision still not well focused; infants rely more on audition, and thus on voice of their caregivers, than on vision to pick up social signals.

\_\_\_-1  
\_\_\_0  
\_\_\_+1

Beyond the discrepancies in the relative timing of the maturation and strength of representation of vocal and facial perception, the evidence presented here underscores the striking similarities between the two processes, which can be related to these evolutionary considerations. These commonalities include early development, progressive refinement as social environments change, and cortical representation becoming more specialized and less diffuse, which conforms with the idea of common organizing principles for representation of face and voice (Yovel & Belin, 2013).

In addition, voice is important as the carrier of speech. The studies we have reviewed often highlight that the maturation of processing of paralinguistic information from voices is not directly related to measures of language maturation (Creel & Jimenez, 2012; Doherty, Fitzsimons, Asenbauer, & Staunton, 1999). Also, children with specific language impairment are not necessarily impaired in tasks that necessitate the processing of nonlinguistic vocal information (Boucher et al., 2000; Levi & Schwartz, 2013). Nevertheless, the development of voice perception interacts with the development of linguistic abilities. As in adults, infants and young children are better at processing language if the speaker's voice is familiar to them (Schmale, Cristia, & Seidl, 2012; van Heugten & Johnson, 2017). In consequence, early development of voice perception skills may help scaffold the building of language.

Conversely, speech recognition may also interact with voice perception development. In adults, bottom-up acoustic analysis and top-down influence from phonological processing collaboratively govern voice recognition (Schmale et al., 2012). Generally we recognize better the identity and emotion, and memorize better a person's voice (Scherer, Banse, & Wallbott, 2001) if she speaks the same language as us, even in conditions when the intelligibility is degraded (Fleming, Giordano, Caldara, & Belin, 2014). This "language familiarity effect" has been described across developmental stages. Infants and preschool-aged children prefer interacting with children and adults of same language (Kinzler, Dupoux, & Spelke, 2007). Seven-month-old infants can distinguish two voices better if the two talkers speak their native language. The strength of the difference in performance in voice recognition for native compared to unfamiliar languages seems to increase during childhood (Levi & Schwartz, 2013). Although this deserves further investigation, these results point toward an intertwining between the development of paralinguistic and linguistic aspects of voice signal processing.

-1\_\_\_  
0\_\_\_  
+1\_\_\_

Thus, the development of voice perception is related to the development of social competencies and to the development of language. This is unlikely to be the whole story, though. A fact that is rarely discussed in studies that tend to focus on averaged age-effect is the large interindividual variability. For instance, in Bartholomeus's study of speaker's recognition in children, the mean performance in four-year-olds was 57 percent correct responses, but it ranged between 11 percent and 95 percent in their sample (Bartholomeus, 1973). In our test of emotion recognition from voice we observed that this interindividual variability decreased from childhood to late adolescence. Although it echoes developmental findings in other domains, the origin of this interindividual variability has yet to be well characterized. It could be related partly to cultural factors (Matsumoto & Kishimoto, 1983).

Interindividual variability is also reflected by the association between voice perception skills and neurodevelopmental pathologies. For instance, children on the autistic spectrum are impaired at recognizing voice identity (Klin, 1991; Schelinski, Riedel, & von Kriegstein, 2014) and emotion (Lindner & Rosen, 2006) whereas the perception of vocal speech is relatively intact, at least when speech is presented with a good signal-to-noise ratio (deGelder, Vroomen, & van der Heide, 1991). In addition, the functional organization of voice-selective cortex in children with autism differs from children with typical development (Yoshimura et al., 2017). Even in children without an autism diagnosis there is a correlation between autistic traits and activity in voice selective cortex (Abrams et al., 2016). Moreover, a deficit in voice perception in adolescents has been associated with socioemotional pathology that emerges in adolescence such as anxiety and depression (McClure & Nowicki, 2001).

## Conclusion

The development of voice processing is special, nonlinear, and relates to social development and communication abilities. It relies on a network of regions in the temporal and frontal cortex, the developmental trajectories of which remain to be better described. Understanding this development in relation to the constraints of different ontogenic stages is important in the study of neurodevelopmental disorders such as autism. In view of the increased use of mediated interactions, especially in the adolescent population, greater understanding is also important for the design of future interacting artificial agents.

\_\_\_-1  
\_\_\_0  
\_\_\_+1



## References

- Abrams, D. A., Chen, T., Odriozola, P., Cheng, K. M., Baker, A. E., Padmanabhan, A., ... Menon, V. (2016). Neural circuits underlying mother's voice perception predict social communication abilities in children. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(22), 6295–6300.
- Allgood, R., & Heaton, P. (2015). Developmental change and cross-domain links in vocal and musical emotion recognition performance in childhood. *British Journal of Developmental Psychology*, *33*(3), 398–403.
- Bakhos, D., Galvin, J., Roux, S., Lescanne, E., & Bruneau, N. (2018). Cortical processing of vocal and nonvocal sounds in cochlear-implanted children: An electrophysiological study. *Ear and Hearing*, *39*(1), 150–160.
- Barker, B. A., & Newman, R. S. (2004). Listen to your mother! The role of talker familiarity in infant streaming. *Cognition*, *94*(2), B45–53.
- Bartholomeus, B. (1973). Voice identification by nursery school children. *Canadian Journal of Psychology*, *27*(4), 464–472.
- Beauchemin, M., Gonzalez-Frankenberger, B., Tremblay, J., Vannasing, P., Martinez-Montes, E., Belin, P., ... Lassonde, M. (2011). Mother and stranger: An electrophysiological study of voice processing in newborns. *Cerebral Cortex*, *21*(8), 1705–1711.
- Belin, P., Fecteau, S., & Bedard, C. (2004). Thinking the voice: Neural correlates of voice perception. *Trends in Cognitive Sciences*, *8*(3), 129–135.
- Belin, P., & Grosbras, M. H. (2010). Before speech: Cerebral voice processing in infants. *Neuron*, *65*(6), 733–735.
- Blasi, A., Mercure, E., Lloyd-Fox, S., Thomson, A., Brammer, M., Sauter, D., ... Murphy, D. G. (2011). Early specialization for voice and emotion processing in the infant brain. *Current Biology*, *21*(14), 1220–1224.
- Bogin, B. (1994). Adolescence in evolutionary perspective. *Acta Paediatrica Supplement*, *406*, 29–35; discussion 36.
- Bonte, M., Frost, M. A., Rutten, S., Ley, A., Formisano, E., & Goebel, R. (2013). Development from childhood to adulthood increases morphological and functional inter-individual variability in the right superior temporal cortex. *Neuroimage*, *83*, 739–750.
- Bonte, M., Ley, A., Scharke, W., & Formisano, E. (2016). Developmental refinement of cortical systems for speech and voice processing. *Neuroimage*, *128*, 373–384.
- Boucher, J., Lewis, V., & Collis, G. M. (2000). Voice processing abilities in children with autism, children with specific language impairments, and young typically developing children. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *41*(7), 847–857.

- Charest, I., Pernet, C. R., Rousselet, G. A., Quinones, I., Latinus, M., Fillion-Bilodeau, S., ... Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience*, *10*, 127.
- Cheng, Y., Lee, S. Y., Chen, H. Y., Wang, P. Y., & Decety, J. (2012). Voice and emotion processing in the human neonatal brain. *Journal of Cognitive Neuroscience*, *24*(6), 1411–1419.
- 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. *British Journal of Developmental Psychology*, *33*(2), 218–236.
- Cohen, M., Prather, A., Town, P., & Hynd, G. (1990). Neurodevelopmental differences in emotional prosody in normal children and children with left and right temporal lobe epilepsy. *Brain and Language*, *38*(1), 122–134.
- Collins, M., & Nowicki, S. (2001). African American children's ability to identify emotion in facial expressions and tones of voice of European Americans. *Journal of Genetic Psychology*, *162*(3), 334–346.
- Creel, S. C., & Jimenez, S. R. (2012). Differences in talker recognition by preschoolers and adults. *Journal of Experimental Child Psychology*, *113*(4), 487–509.
- DeCasper, A. J., & Fifer, W. P. (1980). Of human bonding: Newborns prefer their mothers' voices. *Science*, *208*(4448), 1174–1176.
- deGelder, B., Vroomen, J., & van der Heide, L. (1991). Face recognition and lip-reading in autism. *European Journal of Cognitive Psychology*, *3*(1), 69–86.
- Dehaene-Lambertz, G., Montavont, A., Jobert, A., Alliol, L., Dubois, J., Hertz-Pannier, L., & Dehaene, S. (2010). Language or music, mother or Mozart? Structural and environmental influences on infants' language networks. *Brain and Language*, *114*(2), 53–65.
- Doherty, C. P., Fitzsimons, M., Asenbauer, B., & Staunton, H. (1999). Discrimination of prosody and music by normal children. *European Journal of Neurology*, *6*(2), 221–226.
- Fernald, A. (1993). Approval and disapproval: Infant responsiveness to vocal affect in familiar and unfamiliar languages. *Child Development*, *64*(3), 657–674.
- Fleming, D., Giordano, B., Caldara, R., & Belin, P. (2014). A language-familiarity effect for speaker discrimination without comprehension. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(38), 13795–13798.
- Flom, R., & Bahrick, L. E. (2007). The development of infant discrimination of affect in multimodal and unimodal stimulation: The role of intersensory redundancy. *Developmental Psychology*, *43*(1), 238–252.
- Friederici, A. D. (2005). Neurophysiological markers of early language acquisition: From syllables to sentences. *Trends in Cognitive Science*, *9*(10), 481–488.

- Friend, M. (2000). Developmental changes in sensitivity to vocal paralinguistic cues. *Developmental Science*, 3(2), 148–162.
- Gil, S., Aguert, M., Le Bigot, L., Lacroix, A., & Laval, V. (2014). Children's understanding of others' emotional states: Inferences from extralinguistic or paralinguistic cues? *International Journal of Behavioral Development*, 38(6), 10.
- Grosbras, M. H., Ross, P. D., & Belin, P. (2018). Categorical emotion recognition from voice improves during childhood and adolescence. *Scientific Reports*, 8(1), 14791.
- Grossmann, T., Oberecker, R., Koch, S. P., & Friederici, A. D. (2010). The developmental origins of voice processing in the human brain. *Neuron*, 65(6), 852–858.
- Grossmann, T., Striano, T., & Friederici, A. D. (2005). Infants' electric brain responses to emotional prosody. *NeuroReport*, 16(16), 1825–1828.
- Halit, H., Csibra, G., Volein, A., & Johnson, M. H. (2004). Face-sensitive cortical processing in early infancy. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 45(7), 1228–1234.
- Herba, C., & Phillips, M. (2004). Annotation: Development of facial expression recognition from childhood to adolescence: Behavioural and neurological perspectives. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, 45(7), 1185–1198.
- Joffe, T. H. (1997). Social pressures have selected for an extended juvenile period in primates. *Journal of Human Evolution*, 32(6), 593–605.
- Kinzler, K. D., Dupoux, E., & Spelke, E. S. (2007). The native language of social cognition. *Proceedings of the National Academy of Sciences of the United States of America*, 104(30), 12577–12580.
- Klin, A. (1991). Young autistic children's listening preferences in regard to speech: A possible characterization of the symptom of social withdrawal. *Journal of Autism and Developmental Disorders*, 21(1), 29–42.
- Lee, G. Y., & Kisilevsky, B. S. (2014). Fetuses respond to father's voice but prefer mother's voice after birth. *Developmental Psychobiology*, 56(1), 1–11.
- Levi, S. V., & Schwartz, R. G. (2013). The development of language-specific and language-independent talker processing. *Journal of Speech, Language, and Hearing Research*, 56(3), 913–920.
- Lindner, J. L., & Rosen, L. A. (2006). Decoding of emotion through facial expression, prosody and verbal content in children and adolescents with Asperger's syndrome. *Journal of Autism and Developmental Disorders*, 36(6), 769–777.
- Lloyd-Fox, S., Blasi, A., Mercure, E., Elwell, C., & Johnson, M. (2012). The emergence of cerebral specialization for the human voice over the first months of life. *Social Neuroscience*, 7(3), 317–330.

-1 \_\_\_  
 0 \_\_\_  
 +1 \_\_\_

- Macchi Cassia, V., Pisacane, A., & Gava, L. (2012). No own-age bias in 3-year-old children: More evidence for the role of early experience in building face-processing biases. *Journal of Experimental Child Psychology, 113*(3), 372–382.
- Mann, V. A., Diamond, R., & Carey, S. (1979). Development of voice recognition: Parallels with face recognition. *Journal of Experimental Child Psychology, 27*(1), 153–165.
- Matsumoto, D., & Kishimoto, H. (1983). Developmental characteristics in judgments of emotion from nonverbal vocal cues. *International Journal of Intercultural Relations, 7*(4), 415–424.
- McClure, E. B., & Nowicki Jr, S. (2001). Associations between social anxiety and non-verbal processing skill in preadolescent boys and girls. *Journal of Nonverbal Behavior, 25*(1), 3–19.
- Miller, C. L. (1983). Developmental changes in male/female voice classification by infants. *Infant Behavior and Development, 6*(2–3), 313–330.
- Mills, K. L., Lalonde, F., Clasen, L. S., Giedd, J. N., & Blakemore, S. J. (2014). Developmental changes in the structure of the social brain in late childhood and adolescence. *Social Cognitive and Affective Neuroscience, 9*(1), 123–131.
- Minagawa-Kawai, Y., van der Lely, H., Ramus, F., Sato, Y., Mazuka, R., & Dupoux, E. (2011). Optical brain imaging reveals general auditory and language-specific processing in early infant development. *Cerebral Cortex, 21*(2), 254–261.
- 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.
- Mumme, D. L., Fernald, A., & Herrera, C. (1996). Infants' responses to facial and vocal emotional signals in a social referencing paradigm. *Child Development, 67*(6), 3219–3237.
- Nowicki, S., Jr., & Mitchell, J. (1998). Accuracy in identifying affect in child and adult faces and voices and social competence in preschool children. *Genetic, Social, and General Psychology Monographs, 124*(1), 39–59.
- Poremba, A., Bigelow, J., & Rossi, B. (2013). Processing of communication sounds: Contributions of learning, memory, and experience. *Hearing Research, 305*, 31–44.
- Purhonen, M., Kilpelainen-Lees, R., Valkonen-Korhonen, M., Karhu, J., & Lehtonen, J. (2004). Cerebral processing of mother's voice compared to unfamiliar voice in 4-month-old infants. *International Journal of Psychophysiology, 52*(3), 257–266.
- Rogier, O., Roux, S., Belin, P., Bonnet-Brilhault, F., & Bruneau, N. (2010). An electrophysiological correlate of voice processing in 4- to 5-year-old children. *International Journal of Psychophysiology, 75*(1), 44–47.

Saito, Y., Kondo, T., Aoyama, S., Fukumoto, R., Konishi, N., Nakamura, K., ... Toshima, T. (2007). The function of the frontal lobe in neonates for response to a prosodic voice. *Early Human Development*, 83(4), 225–230.

Sauter, D. A., Panattoni, C., & Happé, F. (2013). Children's recognition of emotions from vocal cues. *British Journal of Developmental Psychology*, 31(1), 97–113.

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.

Scherer, K. R., Banse, R., & Wallbott, H. G. (2001). Emotion inferences from vocal expression correlate across languages and cultures. *Journal of Cross-Cultural Psychology*, 32(1), 76–92.

Scherf, K. S., Behrmann, M., & Dahl, R. E. (2012). Facing changes and changing faces in adolescence: A new model for investigating adolescent-specific interactions between pubertal, brain and behavioral development. *Developmental Cognitive Neuroscience*, 2(2), 199–219.

Schmale, R., Cristia, A., & Seidl, A. (2012). Toddlers recognize words in an unfamiliar accent after brief exposure. *Developmental Science*, 15(6), 732–738.

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.

Stevenage, S. V., Neil, G. J., Barlow, J., Dyson, A., Eaton-Brown, C., & Parsons, B. (2013). The effect of distraction on face and voice recognition. *Psychological Research*, 77(2), 167–175.

Valenza, E., Simion, F., Cassia, V. M., & Umiltà, C. (1996). Face preference at birth. *Journal of Experimental Psychology: Human Perception and Performance*, 22(4), 892–903.

Van Heugten, M., & Johnson, E. K. (2017). Input matters: Multi-accent language exposure affects word form recognition in infancy. *Journal of the Acoustical Society of America*, 142(2), E1196.

Van Lancker, D., Cornelius, C., & Kreiman, J. (1989). Recognition of emotional-prosodic meanings in speech by autistic, schizophrenic, and normal children. *Developmental Neuropsychology*, 5(2–3), 207–226.

Vouloumanos, A., Hauser, M. D., Werker, J. F., & Martin, A. (2010). The tuning of human neonates' preference for speech. *Child Development*, 81(2), 517–527.

Wade, A. M., Lawrence, K., Mandy, W., & Skuse, D. (2006). Charting the development of emotion recognition from 6 years of age. *Journal of Applied Statistics*, 33(3), 297–315.

Walker-Andrews, A. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin*, 121(3), 437–456.

-1  
0  
+1

Walker-Andrews, A., & Lennon, E. (1991). Infants' discrimination of vocal expressions: Contributions of auditory and visual information. *Infant Behavior and Development*, *14*, 131–142.

Yamasaki, T., Ogata, K., Maekawa, T., Ijichi, I., Katagiri, M., Mitsudo, T., ... Tobimatsu, S. (2013). Rapid maturation of voice and linguistic processing systems in preschool children: A near-infrared spectroscopic study. *Experimental Neurology*, *250*, 313–320.

Yoshimura, Y., Kikuchi, M., Hayashi, N., Hiraishi, H., Hasegawa, C., Takahashi, T., ... Minabe, Y. (2017). Altered human voice processing in the frontal cortex and a developmental language delay in 3- to 5-year-old children with autism spectrum disorder. *Scientific Reports*, *7*(1), 17116.

Yovel, G., & Belin, P. (2013). A unified coding strategy for processing faces and voices. *Trends in Cognitive Sciences*, *17*(6), 263–271.

Zhang, D., Zhou, Y., Hou, X., Cui, Y., & Zhou, C. (2017). Discrimination of emotional prosodies in human neonates: A pilot fNIRS study. *Neuroscience Letters*, *658*, 62–66.

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.

-1\_\_  
0\_\_  
+1\_\_