

Sex differences in the recognition of emotional prosody in late childhood and adolescence

Takashi X. Fujisawa · Kazuyuki Shinohara

Received: 29 December 2010 / Accepted: 22 May 2011 / Published online: 7 June 2011
© The Physiological Society of Japan and Springer 2011

Abstract We examined sex-related differences in the ability to recognize emotional prosody in late childhood (9–12 year olds) and adolescence (13–15 year olds) in relation to salivary testosterone levels. In order to examine both the accuracy and the sensitivity in labeling emotional prosody expressions, five intensities (20, 40, 60, 80, and 100%) for each of three emotion categories were used as stimuli. Totals of 25 male and 22 female children and 28 male and 28 female adolescents were tested on their recognition of happy, angry and sad prosody at the different intensities. The results showed that adolescent females were more sensitive to happy and sad prosody than males but not to angry prosody, whereas there were no sex-related differences in emotional prosody in late childhood for any of the emotional categories. Furthermore, salivary testosterone levels were higher in males than females in adolescence, but not in late childhood, suggesting that the sex differences for emotional prosody recognition emerges in adolescence during which testosterone levels become higher in males than females.

Keywords Emotion · Prosody · Recognition · Sex difference · Puberty

Introduction

We need to pay attention to the emotions of others in order to successfully create human relationships since

emotional signals help us to predict the actions of others and to adjust our own behavior appropriately. Besides the semantic meaning of words, emotional prosody in speech may convey important information about the emotional state of the speaker. Emotional prosody is defined as the ability to express emotions through various acoustical parameters of human speech, such as pitch, intensity and duration [1]. A specific profile of these parameters can be described by a specific emotion. For example, anger is typical of a relatively high intensity, speech rate and mean fundamental frequency (F0), whereas sadness is typical of low intensity, speech rate and mean F0 [2]. This signaling function plays a significant role in everyday communication and has the potential to determine the meanings of a contradicting verbal message (e.g., irony).

It is the general consensus that women recognize emotional expression from faces [3], body movements [4] and voices [5] more accurately than do men. While many behavioral studies have reported sex-related differences in the perception of facial emotions (see [3] for a review), only a few studies have paid attention to the processing of emotional information in the voice [6]. Behavioral studies on prosody have focused on sex differences in phonological rather than emotional tasks [7–10]. Recently, several laboratories have developed lines of research that investigate emotional prosody processing in men and women. Schirmer and colleagues [11], in a cross-modal priming experiment, suggesting that women may make faster use of emotional prosody during language processing. Everhart and colleagues [5] examined the processing of emotional utterances to avoid potential confounds with linguistic processing. The sex-related main effect also suggests that women may process emotional prosody more quickly than do men.

T. X. Fujisawa · K. Shinohara (✉)
Department of Neurobiology and Behavior,
Nagasaki University School of Medicine,
1-12-4 Sakamoto, Nagasaki 852-8523, Japan
e-mail: kazuyuki@nagasaki-u.ac.jp

When do the sex-related differences in emotional prosody emerge in life? Sexual differentiation for emotional prosody perception may result from sexual differentiation for structural correlates of the function. Many sex differences in brain structures seem to occur after the onset of puberty (see [12] for a review) although developmental trajectories differ among structures. Especially, the temporal lobe including regions associated with emotional prosody [13, 14] continues to develop structurally throughout late childhood and adolescence [15, 16].

In light of the above findings, we have investigated sex differences in the recognition ability of emotional prosody in late childhood and adolescence. With regard to emotions, we focused on happy, sad and angry emotions because a previous adult study investigated sex differences in happy, sad and angry prosody recognition [6]. In the present study, we used blends of happy, sad and angry with neutral prosody expressions to determine the overall accuracy and sensitivity of each group to the target emotion.

Sexual differentiation for emotional prosody and its responsible structures should be affected by hormonal maturation more strongly than chronological age [17], so we have measured salivary levels of testosterone in late childhood and adolescence.

Methods

Participants

The participants were recruited from two elementary schools and one junior high school in two geographical areas of Nagasaki prefecture in Japan, representative of a range of socioeconomic status (SES) backgrounds. Exclusion criteria for children and adolescents were: (1) current or lifetime history of neurological, or psychiatric disorder, (2) current or lifetime history of hearing loss, and (3) full scale IQ at 5 years of age lower than 80. The experimental protocol was conducted according to the tenets of the Helsinki Declaration and was approved by the Ethics Committee of the Nagasaki University.

The late childhood group consisted of 22 female (mean age 10.1 ± 1.24 years) and 25 male (mean age 10.4 ± 1.65 years) children aged 9–12 years (mean age 10.3 ± 1.46 years). The adolescent group consisted of 28 female (mean age 14.4 ± 0.93 years) and 28 male adolescents (mean age 14.4 ± 0.68 years) aged 13–15 years (mean age 14.4 ± 0.82 years). Mean ages for females and males were not significantly different for each age group. In the present study, we investigated these groups (late childhood and adolescence) as distinct experimental groups for the following reasons: first, to control the socio-

environmental factor between the groups because they were at different stages in the Japanese educational curriculum; second, to control the factor of cognitive functions because it has been suggested [18] that Japanese children at the age of about 12 or 13 are in a state of cognitive transition; and third, they were separated into two groups to control for factors related to psychological development because, according to the traditional theory of developmental psychology [19], they are likely to be at different stages.

Measures

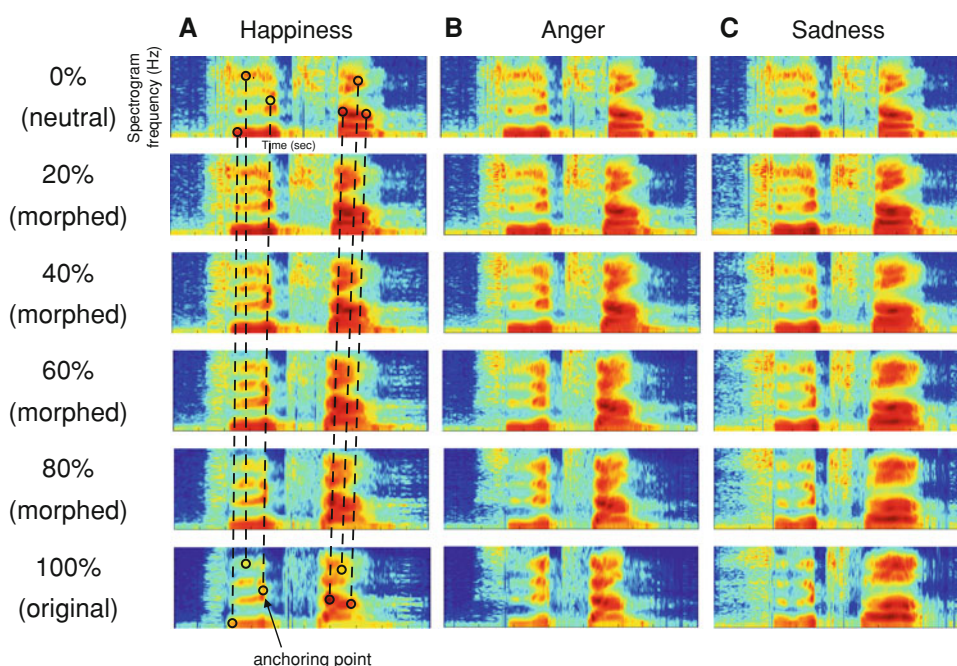
Hormone assays

Saliva samples were collected from each participant between 1200 and 1300 hours on the day of the experiment (15–30 min before the noon meal), by having them spit through a straw into a small polystyrene tube. Saliva samples were frozen and stored at -80°C in the laboratory. Testosterone was assayed in saliva duplicates using an ELISA technique (Salimetrics, State College, USA), with each sample being analyzed in duplicate. The average intra-assay coefficient of variation (CV) was 4.1%.

Emotion recognition of vocal expressions

These vocal stimuli were taken from our original dataset where four professional actors (two boys, two girls) produced a one-word utterance “Suzuki-san” (the most popular name in Japan) with happy, angry, sad and neutral prosody for each age group (late childhood and adolescence). Voices were recorded in a sound-proof room of a motion picture studio, using a MKH416 condenser microphone (Sennheiser) at a distance of approximately 30 cm, and fed into a digital audiotape recorder (Sony, TCD-10PRO) with 16-bit resolution and a sampling rate of 48 kHz. They were then edited into meaningful segments with a normalized peak value (90% of maximum amplitude) and were downsampled at 44.1 kHz, using SoundEngine Free (Cycle 5th). For each actor and vocalization category, only the best occurrence, evaluated as successful by the two experimenters, was included in the dataset. For each emotion, these sounds were morphed at 6 intensity levels (0% as an emotionally neutral stimulus, and at 20, 40, 60, 80 and 100% for stimuli with emotional valence) using STRAIGHT software [20]. Here, the 100% intensity level was the original voice recording uttered by actors with each emotion. STRAIGHT is a high-quality speech modification procedure and can be used to enable auditory morphing for CD quality test stimuli. Morphing voices using the STRAIGHT technique is very simple, and can be executed interactively using Matlab GUI. First, the F0 and aperiodicity were extracted from the target two

Fig. 1 Spectrograms of voice samples of the three emotion morph continua used in this experiment: **a** neutral to happy, **b** neutral to angry, and **c** neutral to sad



voices for morphing, and then the STRAGHT spectrogram calculation was executed. After that, the anchoring points for the target voices using the prepared GUI were set, and a morphed sound (similar to morphing two faces using suitable software) was synthesized (Fig. 1).

The experiments were conducted in a quiet room of the participants' school. Voice stimuli were presented for 1.5 s via speakers connected to a computer placed in front of the subjects. The level of the audio signal was adjusted to clearly audible. The stimulus presentation and data collection were controlled by a PC running SuperLab software (Cedrus, San Pedro, USA). There were 64 stimuli overall [(5 intensity levels $\times$ 3 expressions $\times$ 4 actors) + (neutral expression $\times$ 4 actors)], presented in random order in two blocks of 64 (each stimulus was presented once in each block). Participants were asked to categorize sounds into four types (happiness, anger, sadness and neutral) following a schematic face presented on the computer screen (Fig. 2). Ratings were obtained after each presentation using the computer keyboard.

Accuracy

Figure 3 shows the mean accuracy at each intensity level for each expression when the data were first averaged across the four models seen by each participant and then averaged across participants in each age group. Most errors occurred at low intensities of the morphed emotions, where expressive voices were identified as neutral. We measured these errors by calculating the threshold to recognize each expression as different from neutral.

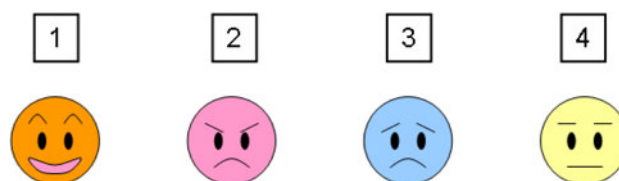


Fig. 2 Schematic faces. From left to right: happy, angry, sad, and neutral

Sensitivity (thresholds)

To measure the children's sensitivity to vocal expressions, we calculated their thresholds to differentiate each vocal expression from neutral. The sensitivity of each participant represents the recognizable tendency for the whole category of the target emotion which is independent from the morph levels. Responses by each participant were categorized as neutral or nonneutral, with nonneutral responses including both correct recognitions (e.g., 60% anger classified as angry) and incorrect recognitions (e.g., 60% anger classified as happy). With an increment of morph levels, the percentage of the correct responses should show a corresponding trend. We fitted a linear function to the percentage of the correct responses of each participant for each expression. We estimated two parameters: *a*, the slope, and *b*, the intercept. The estimated value of *b* was used as the threshold and corresponds to $p = .5$, where p is the probability of recognition. In other words, the threshold value means the intensity level at which the expressive voice will be recognized as neutral 50% of the time and recognized as

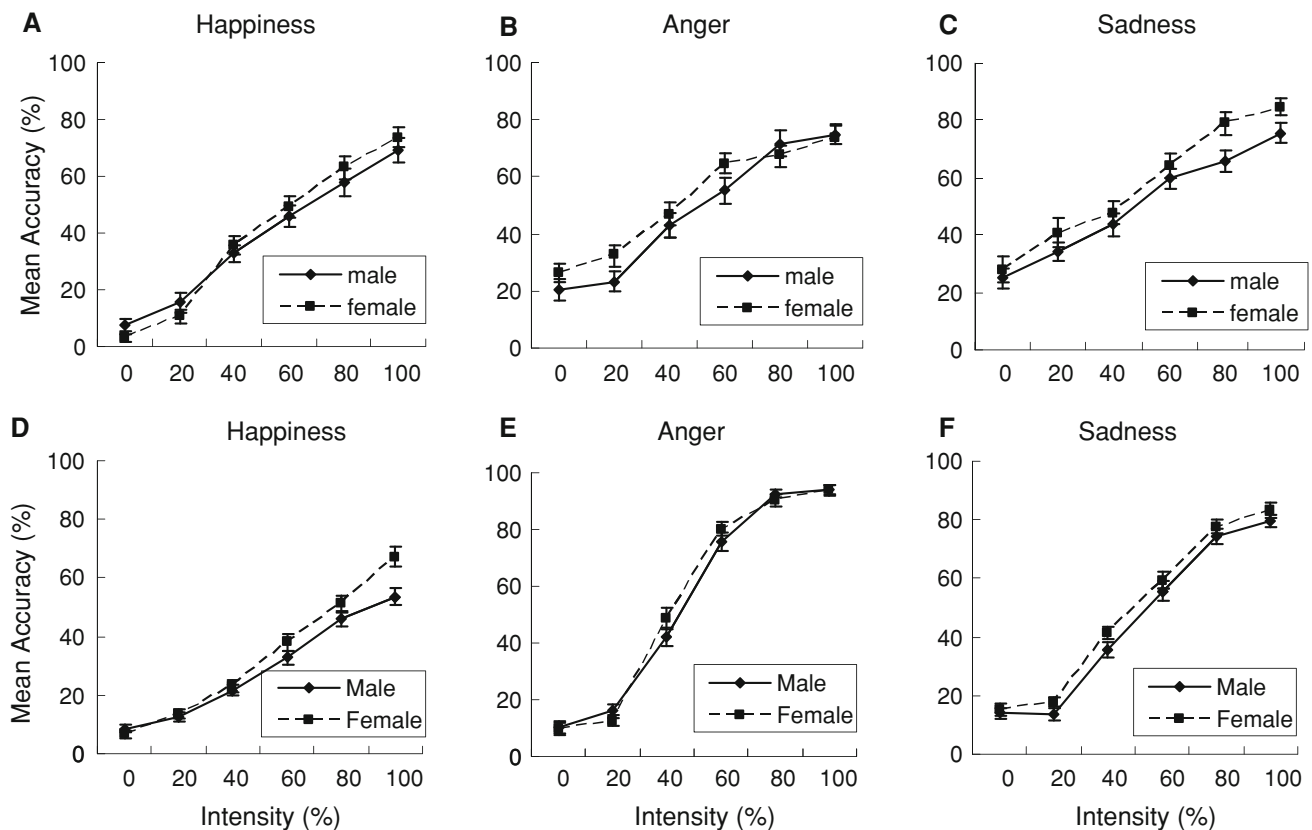


Fig. 3 Mean accuracy (± 1 SE) for each expression in both age groups (**a–c** late childhood, **d–f** adolescence)

corresponding to one of the three emotional expressions in the remaining 50%. The individual's threshold for each expression was averaged across the independently derived estimates for the four models.

Results

Hormone levels

Table 1 shows salivary testosterone levels for the male and female subjects for each age group. To investigate the effect of sex for salivary testosterone level, 2×2 ANOVAs (sex $\times$ age group) were conducted, sex (male, female) and age group (late childhood, adolescence) as between-subject factors. First, with regard to the testosterone level, the interaction of sex $\times$ age group were significant [$F(5,270) = 4.01$, $p < .005$]. The test of post hoc comparison revealed that a simple main effect of sex was not significant in late childhood group ($p > .10$), but was significantly different in the adolescent group ($p < .001$). These results suggest that the mean testosterone level for males was higher than for females in adolescence, while gender differences were not observed in late childhood.

Accuracy

Figure 3 shows the accuracy for the separate emotions for both groups. To investigate the effect of sex, 2×6 ANOVAs (sex $\times$ intensity level) were conducted for each emotion, sex (male, female) as between-subject variable and intensity level (6 levels: 0, 20, 40, 60, 80, 100%) as within-subject factors. First, for the happy voice, the main effect of sex and the interaction of sex $\times$ intensity level were not significant in the late childhood group [$F(1,39) = 0.17$, n.s.; $F(5,195) = 0.95$, n.s., respectively], but in the adolescent group, sex $\times$ intensity level interaction was found [$F(5,270) = 4.01$, $p < .005$]. The test of post hoc comparison revealed a simple main effect of sex [$F(1,324) = 17.89$, $p < .001$] for the 100% intensity level. These results suggest that female adolescents were more accurate in recognition of the happy voice with strong expression than males. Next, for the angry voice, the main effect of sex and the interaction of sex $\times$ intensity level were not significant in the late childhood group [$F(1,42) = 1.10$, n.s.; $F(5,210) = 1.48$, n.s., respectively]. Similarly in the adolescent group, the main effect and the interaction were not significant [$F(1,54) = 0.01$, n.s.; $F(5,270) = 1.71$, n.s., respectively]. These results suggest that there is no sex difference of accuracy in recognition of

the angry voice in these developmental stages. Finally, for the sad voice, the main effect of sex and the interaction of sex $\times$ intensity level were not significant in the late childhood group [$F(1,42) = 2.56$, n.s.; $F(5,210) = 0.85$, n.s., respectively], but in the adolescent group, the main effect of sex was significant [$F(1,54) = 5.62$, $p < .05$]. These results suggest that female adolescents were more accurate in recognition of the sad voice than males.

Sensitivity

For the threshold rates, we tested the effect of sex difference separately for each expression for each age group (Fig. 4). First, for the happy voice, the sensitivity score was not significant in the late childhood group [$t(38) = 1.66$, n.s.], but was significantly different in the adolescent group [$t(54) = 1.93$, $p < .05$]. These results suggest that female adolescents were more sensitive in recognition of the happy voice than males, while there was no sex difference of sensitivity in recognition of the happy voice in both age groups. Next, for the angry voice, the sensitivity score was not significantly different in common to both age groups [$t(40) = 1.34$, n.s.; $t(54) = 0.01$, n.s.; respectively]. Finally, for the sad voice, the sensitivity score was not significant in the late childhood group [$t(41) = 1.42$, n.s.], but in the adolescent group, the score was significantly different [$t(54) = 3.76$, $p < .001$]. These results suggest that female adolescents were more sensitive in recognition of the sad voice than males, while there was no sex

differences of sensitivity in recognition of the sad voice in the late childhood group.

Discussion

The current study evaluated the sex difference for prosody recognition abilities in late childhood and adolescence by means of morphing techniques. The results revealed that adolescent females were more sensitive to happy and sad prosody than males, whereas these differences were not shown in late childhood for any emotional category of prosody. On the other hand, salivary testosterone levels were higher in males than females in adolescence, but not in late childhood. These findings suggest that the sex differences for emotional prosody recognition emerge in adolescence during which testosterone levels become higher in males than females.

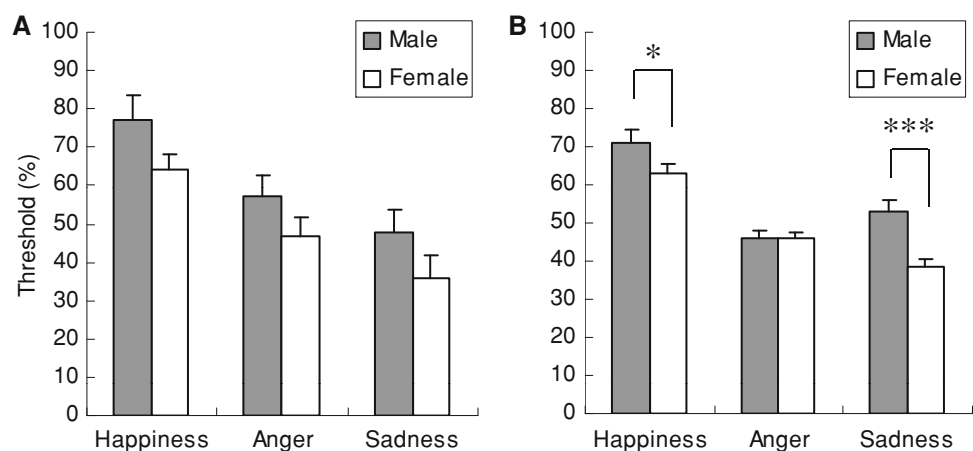
The present study investigated the influence of the actual physical intensity of a particular expression on decoding accuracy for emotional speech prosody. For this purpose, we used a new technique, STRAIGHT [20], which enabled us to manipulate speech parameters to vary the emotional intensity. The validity of this procedure has been demonstrated by previous studies showing that the naturalness of morphed speech samples was comparable to natural speech samples, and that this procedure allows for the creation of a continuum of stimuli that vary gradually between different emotional expressions [21]. Using this technique, we found sex differences in the recognition for sad expressions regardless of the intensity, and for happy expressions only when the emotional intensity was strong. To our knowledge, this is the first study to investigate the relationship between the expressional intensity of emotional prosody and decoding accuracy for each emotional category.

One of the most interesting findings in the present study is sex differences in recognition ability for emotional prosody in adolescence, around 14 years of age. Adolescent

Table 1 Means, standard deviations, and test scores for each group

	Salivary testosterone (pg/ μ l)		<i>p</i> value
	Male	Female	
Late childhood	32.0 (25.1)	39.7 (27.5)	n.s.
Adolescence	107.5 (47.5)	49.2 (22.8)	<.001

Fig. 4 Mean sensitivity (± 1 SE) for each expression in both age groups (**a** late childhood, **b** adolescence)



females were more sensitive to happy and sad prosody than males. These results were consistent with a previous adult study, which demonstrated that females judged happy and sad prosody more accurately than males [6]. Similarly, our current results were in line with electrophysiological studies, showing better accuracy in emotional prosody recognition [22, 23] and faster recognition and enhanced cortical responses in adult females than males [24]. Regarding the superior sensitivity of females for emotional expression, consistent results were reported for facial expressions [25, 26] and body movement studies [4].

On the other hand, a sex difference in sensitivity to emotional prosody was observed only in adolescence, but not in late childhood. Adolescence is an important developmental period in which major physical, psychological, cognitive, and social transformations occur [27–29], and gender differences emerge and manifest themselves [30, 31]. Hence, differences in sensitivity to emotional prosody are likely. Indeed, neurodevelopmental studies suggest that gray matter (GM) volume of the temporal lobe, part of which is involved in emotional prosody perception [14, 32, 33], continues to develop structurally throughout late childhood and adolescence in a gender-dependent fashion [15, 16]. In contrast, the GM volume of the frontal and parietal lobes matures much earlier in late childhood [15].

The emergence of sex differences in prosody recognition during adolescence is likely consequent on steroid-dependent activation and/or organization of brain structures associated with prosody perception. Several lines of evidence suggest that androgen receptors (AR) are abundant in the human temporal cortex [34, 35]. In the current study, higher levels of testosterone were observed in adolescent males than females. Additionally, the negative relationship between recognition accuracy and testosterone level was also found within the adolescent male group in our preliminary investigation. Moreover, although we did not investigate the sex difference in adults, a female superiority in sensitivity for emotional prosody was found [6]. Taking these findings together, the adolescent increases in testosterone might produce sex differentiation for activation and/or organization of the male temporal lobe, which plays an important role in detecting emotional prosody. However, because females start biological maturation at puberty about 2 years earlier than males [36], it cannot be ruled out that the higher sensitivity to emotional prosody in adolescent females was caused by factors of biological development. This point clearly indicates a limitation in the present study, and further investigation is needed.

In conclusion, the present results show a higher sensitivity for happy and sad prosody in adolescent females than males while there was no difference in sensitivity to angry prosody in either sex. These sex differences were not

observed in late childhood. Moreover, we pointed to the possibility that these differences were modulated by a steep increase in testosterone levels in adolescence. In the present study, we could not clarify the neural mechanism underlying this effect. Combining the present experimental paradigm with neurophysiological indicators of temporal cortex activities such as EEG and near-infrared spectroscopy (NIRS) will be fruitful in further elucidating the mechanisms underlying sexual differentiation for human prosody perception.

Acknowledgments The authors would like to thank the students and school teachers, especially Hidenori Ejima, Yukari Mori and Katsuki Fukuda, who made this research possible.

References

1. Besson M, Magne C, Schön D (2002) Emotional prosody: sex differences in sensitivity to speech melody. *Trends Cogn Sci* 6(10):405–407
2. Banse R, Scherer KR (1996) Acoustic profiles in vocal emotion expression. *J Pers Soc Psychol* 70(3):614–636
3. McClure EB (2000) A meta-analytic review of sex differences in facial expression processing and their development in infants, children, and adolescents. *Psychol Bull* 126(3):424–453
4. Sogon S, Izard CE (1987) Sex differences in emotion recognition by observing body movements: a case of American students. *Jpn Psychol Res* 29(2):89–93
5. Everhart DE, Demaree HA, Shipley AJ (2006) Perception of emotional prosody: moving toward a model that incorporates sex-related differences. *Behav Cogn Neurosci Rev* 5(2):92–102
6. Bonebright TL, Thompson JL, Leger DW (1996) Gender stereotypes in the expression and perception of vocal affect. *Sex Roles* 34(5–6):429–445
7. Pugh KR, Shaywitz BA, Shaywitz SE, Constable RT, Skudlarski P, Fulbright RK et al (1996) Cerebral organization of component processes in reading. *Brain* 119(Pt 4):1221–1238
8. Pugh KR, Shaywitz BA, Shaywitz SE, Shankweiler DP, Katz L, Fletcher JM et al (1997) Predicting reading performance from neuroimaging profiles: the cerebral basis of phonological effects in printed word identification. *J Exp Psychol* 23(2):299–318
9. Shaywitz BA, Shaywitz SE, Pugh KR, Constable RT, Skudlarski P, Fulbright RK et al (1995) Sex differences in the functional organization of the brain for language. *Nature* 373(6515):607–609
10. Plante E, Schmithorst VJ, Holland SK, Byars AW (2006) Sex differences in the activation of language cortex during childhood. *Neuropsychologia* 44(7):1210–1221
11. Schirmer A, Kotz SA, Friederici AD (2002) Sex differentiates the role of emotional prosody during word processing. *Cogn Brain Res* 14(2):228–233
12. Lenroot RK, Giedd JN (2006) Brain development in children and adolescents: insights from anatomical magnetic resonance imaging. *Neurosci Biobehav Rev* 30(6):718–729
13. Grandjean D, Sander D, Pourtois G, Schwartz S, Seghier ML, Scherer KR, Vuilleumier P (2005) The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat Neurosci* 8(2):145–146
14. Ethofer T, Anders S, Wiethoff S, Erb M, Herbert C, Saur R, Grodd W, Wildgruber D (2006) Effects of prosodic emotional intensity on activation of associative auditory cortex. *Neuroreport* 17(3):249–253

15. Giedd JN, Blumenthal J, Jeffries NO, Castellanos FX, Liu H, Zijdenbos A, Paus T, Evans AC, Rapoport JL (1999) Brain development during childhood and adolescence: a longitudinal MRI study. *Nat Neurosci* 2(10):861–863
16. Bramen JE, Hranilovich JA, Dahl RE, Forbes EE, Chen J, Toga AW, Dinov ID, Worthman CM, Sowell ER (2011) Puberty influences medial temporal lobe and cortical gray matter maturation differently in boys than girls matched for sexual maturity. *Cereb Cortex* 21(3):636–646
17. Neufang S, Specht K, Hausmann M, Güntürkün O, Herpertz-Dahlmann B, Fink GR, Konrad K (2009) Sex differences and the impact of steroid hormones on the developing human brain. *Cereb Cortex* 19(2):464–473
18. Mizuno K, Tanaka M, Fukuda S, Sasabe T, Imai-Matsumura K, Watanabe Y (2011) Changes in cognitive functions of students in the transitional period from elementary school to junior high school. *Brain Dev* 33(5):412–420
19. Erikson EH (1959) *Identity and the life cycle*. International Universities Press, New York
20. Kawahara H, Masuda-Kasuse I, de Cheveigne A (1999) Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: possible role of a repetitive structure in sounds. *Speech Commun* 27(3–4):187–207
21. Matsui H, Kawahara H (2003) Investigation of emotionally morphed speech perception and its structure using a high quality speech manipulation system. *Proc Eurospeech 2003*:2113–2116
22. Hall JA (1978) Gender effects in decoding nonverbal cues. *Psychol Bull* 85(4):845–857
23. Belin P, Fillion-Bilodeau S, Gosselin F (2008) The Montreal Affective Voices: a validated set of nonverbal affect bursts for research on auditory affective processing. *Behav Res Methods* 40(2):531–539
24. Schirmer A, Kotz SA (2006) Beyond the right hemisphere: brain mechanisms mediating vocal emotional processing. *Trends Cogn Sci* 10(1):24–30
25. Kirouac G, Dore EY (1985) Accuracy of the judgments of facial expression of emotions as a function of sex and level of education. *J Nonverbal Behav* 9(1):3–7
26. Wallbott HG (1988) Big girls don't frown, big boys don't cry: gender differences of professional actors in communicating emotion via facial expression. *J Nonverbal Behav* 12(2):98–106
27. Blakemore S-J, Choudhury S (2006) Development of the adolescent brain: implications for executive function and social cognition. *J Child Psychol Psychiatry* 47(3–4):296–312
28. Galvan A, Hare T, Voss H, Glover G, Casey BJ (2007) Risk-taking in the adolescent brain: who is at risk? *Dev Sci* 10(2):F8–F14
29. Markham JA, Morris JR, Juraska JM (2007) Neuron number decreases in the rat ventral, but not dorsal, medial prefrontal cortex between adolescence and adulthood. *Neuroscience* 144(3):961–968
30. Lenroot RK, Gogtay N, Greenstein DK, Wells EM, Wallace GL, Clasen LS, Blumenthal JD, Lerch J, Zijdenbos AP, Evans AC et al (2007) Sexual dimorphism of brain development trajectories during childhood and adolescence. *Neuroimage* 36(4):1065–1073
31. Schmithorst VJ, Holland SK, Dardzinski BJ (2007) Developmental differences in white matter architecture between boys and girls. *Hum Brain Mapp* 29(6):696–710
32. Grandjean D, Sander D, Pourtois G et al (2005) The voices of wrath: brain responses to angry prosody in meaningless speech. *Nat Neurosci* 8(2):145–146
33. Beaucois V, Lacheret A, Turbelin MR, Morel M, Mazoyer B, Tzourio-Mazoyer N (2007) FMRI study of emotional speech comprehension. *Cereb Cortex* 17(2):339–352
34. Sarrieau A, Mitchell JB, Lal S, Olivier A, Quirion R, Meaney MJ (1990) Androgen binding sites in human temporal cortex. *Neuroendocrinology* 51(6):713–716
35. Puy L, MacLusky NJ, Becker L, Karsan N, Trachtenberg J, Brown TJ (1995) Immunocytochemical detection of androgen receptor in human temporal cortex. *J Steroid Biochem Mol Biol* 55(2):197–209
36. Tanner JM, Whitehouse RH (1976) Clinical longitudinal standards for height, weight, height velocity, weight velocity, and stages of puberty. *Arch Dis Child* 51(3):170–179