Speech Communication 183 (2026) 103452

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Acoustic features underlying internal representation of vocal smile in children

Zoé Ranty [a], Annabelle Merchie [b], Jean-Julien Aucouturier [c], Marie Gomot [a,*]

[a] Université de Tours, INSERM, Imaging Brain & Neuropsychiatry iBraiN U1253, 37032, Tours, France
[b] ACTE at ULB Neuroscience Institute, Université Libre de Bruxelles, 50 Avenue F.D. Roosevelt, Brussels, 1050, Belgium
[c] Université Marie et Louis Pasteur, SUPMICROTECH, CNRS, institut FEMTO-ST, F-25000 Besançon, France

ARTICLE INFO

Keywords:
Emotion
Development
Vocal smile
Reverse correlation
Emotional prosody
Psychophysics
Acoustics

ABSTRACT

Smile is a key feature in development and can also be heard in the voice. During development, an improvement of vocal emotion recognition abilities has been described along with global neuroanatomical maturation of auditory cortex. While it has been suggested by the predictive coding framework that internal representation refines with age during development, internal representation of vocal smile has not been characterized in children within this specific age range. The present study aims to bridge this knowledge gap by investigating internal representation of vocal smile in children, along with the robustness of this representation.

Vocal smile model was assessed in 46 children aged 8 to 12 years and 42 adults. A reverse correlation approach was implemented allowing to distinguish two components of vocal smile model: the acoustic features underlying a vocal smile and the internal representation robustness, quantified as internal Noise.

As a group, children display the same internal representation of vocal smile as adults. However, when comparing children's individual representations to adults one, we demonstrated that within children group, their representation becomes increasingly closer to adults' template with age. Furthermore, internal noise estimation allows to distinguish two children subgroup, with one subgroup exhibiting a delay in the acquisition of that internal representation.

This study provides key insights into the development and variability of emotional voice processing in school-age children.

1. Introduction

The ability to recognize emotions is considered as a crucial life skill, particularly during early childhood (Herba and Phillips, 2004). Children who excel at recognizing others' emotions tend to perform better academically (Widen, 2020; Torres et al., 2015) demonstrate greater self-confidence (Widen, 2020) and show improved social adjustment (Goodfellow and Nowicki Jr., 2009). Among the various modalities of emotion recognition, the ability to perceive and interpret emotional cues in speech, known as emotional prosody, plays a fundamental role in social and cognitive development.

1.1. Recognition of emotional prosody in children

Emotional prosody refers to the modulation of multiple acoustic cues (pitch, intensity, speech rate and timbre) to convey emotion through voice. Because hearing becomes functional around the 20th week of gestation, exposure to spoken language begins in utero (Eggermont and Moore, 2012; Pujol et al., 1991). Recent evidence indicates a developmental shift in emotional prosody discrimination at approximately 37 weeks of gestational age, with neonates born after this period exhibiting a greater ability to distinguish between happy and neutral vocal expressions than those born between 35 and 37 weeks (Hou et al., 2024). From the age of 2 months, infants show a higher level of attention to emotional speech addressed to them (parenthese or Infant Directed Speech). Parenthese relies on acoustic features closer to happy prosody, than to neutral speech addressed to adults (Pegg et al., 1992). Thus, prosody is one of the main aspects of language to be processed by the baby, enabling voice communication without verbal semantic content.

Across typical development, it is well established that the ability to

* Corresponding author at: Université de Tours, Inserm, iBraiN U1253, Centre de Pédopsychiatrie, CHRU Bretonneau 2 Boulevard Tonnellé, 37044 Tours, Cédex 09, France.
E-mail address: gomot@univ-tours.fr (M. Gomot).

https://doi.org/10.1016/j.specom.2026.103452
Received 21 October 2025; Received in revised form 1 July 2026; Accepted 15 July 2026
Available online 16 July 2026

recognize emotions in general, and emotional prosody in particular, improves with age (Riddell et al., 2024; Amorim et al., 2021). By seven months of age, exposure to semantically neutral words spoken with emotional prosody elicits distinct cerebral responses (Grossmann et al., 2005). Behaviorally, however, children aged 2–3 years remain unable to use prosodic contours to infer the emotion being expressed, relying instead on bodily cues. From around 4–5 years of age, children can interpret the exaggerated prosodic contours associated with joy (Quam and Swingley, 2012), yet they still do not prioritize prosody over content (Golan et al., 2007), even when visual cues reinforce the conveyed emotion (Filippi et al., 2017). According to Matsumoto & Lee (1993), children begin to accurately label basic emotions in a speaker's voice between 6 and 9 years of age. In children aged 5–10 years, Sauter and colleagues (2013) reported that age predicts overall emotion recognition performance for both speech stimuli (affectively inflected spoken three-digit numbers) and nonverbal vocalizations (e.g., laughs, grunts, sighs), with even 5-year-olds performing above chance. Similarly, Aguert et al. (2013) found that emotion recognition from speech with unintelligible lexical content but salient emotional prosody remained at chance level in 5-year-olds, but improved significantly in children aged 9–13 years.

This critical window for the development of emotional prosody recognition may be associated with the anatomical maturation of the auditory cortex, which is also reported to occur in this specific age-range. Auditory cortical maturation is thought to originate within superficial cortical layers II and III, driven by neurodevelopmental processes such as neurofilament proliferation and axonal myelination (Eggermont and Moore, 2012; Moore and Guan, 2001). These neurophysiological modifications are also thought to be reflected in the morphological changes observed in Auditory Evoked Potentials (Ponton et al., 2000)

Beyond this early developmental period, emotion recognition abilities continue to improve throughout adolescence. Between 6 and 17 years of age, the ability to recognize target emotions in pseudowords spoken with different emotional intonations by adults increases progressively with age (Filippa et al., 2022), although performance does not yet reach adult levels (Aguert et al., 2013). In adulthood, emotional prosody recognition appears largely universal across cultures and languages (Sauter et al., 2010; Pell et al., 2009; Bryant and Barrett, 2007). However, recognition performance tends to decline in older adults, resulting in an overall inverted U-shaped developmental trajectory (Amorim et al., 2021)

Among the various forms of emotional prosody, one particularly intriguing aspect is the vocal smile— the sound produced when a speaker smiles while speaking. Smiling, in general, represents an important evolutionary feature, serving as a means of conveying and perceiving positive emotions, thereby facilitating social interaction (Ekman, 1992; Eibl-Eibesfeldt, 2017). Although the majority of research on smiling has focused on its facial expression (e.g., Rychlowska, 2017), smiling is not only perceived visually but can also be detected acoustically in the human voice. For instance, the acoustic consequences of smiling on pitch, intensity and timbre during speech production have been studied using corpora of recorded speech (Basso and Oullier, 2010; Barthel and Quené, 2015; Podesva et al., 2015). El Haddad et al. (2015) recorded a speaker reading phonetically balanced sentences with either speech-smile or neutral speech to analyze the corresponding acoustic features. In comparison to neutral speech, first and second formants increase of speech-smile was similar to speech-laugh and laugh (Menezes and Igarashi, 2006). These studies suggest that such acoustic modulations correspond to well-established internal representation in adults. In particular, adults rely on a consistent set of spectral cues to identify a smiling voice—findings that have proven remarkably robust across studies (Ponsot et al., 2018a; Merchie et al., 2024). Specifically, smiling speech is characterized by increased intensity and elevated frequencies in the first formant (the peak of the spectral envelope), as well as higher frequencies in the second and third formants.

1.2. Internal representations of emotion

A simple yet influential framework for explaining the perceptual and cognitive processing of sensory cues such as emotional prosody, is based on the concept of internal "templates", or "representations". Inspired by classical observer models in psychophysics this approach posits that listeners recognize specific vocal expressions by comparing incoming stimuli with pre-learned internal representations, although the exact mechanisms underlying this comparison may vary. For instance, in the linear observer model, stimuli are compared with a pre-learned template, via a simple scalar product, later summed with a gaussian, additive "internal noise" to model response inconsistencies (Murray, 2011)

While theory abounds in how such internal representation may be used by adult observers (for a review, see Murray, 2011) the question of how children may develop such psychophysical templates, particularly in the context of emotion recognition, remains remarkably underexplored. Within the framework of predictive coding, it has been proposed that the infant brain possesses a set of inherited expectations about the world, often referred to as innate priors (Badcock et al., 2019). As described in a recent review (Rapaport and Sowman, 2024), these innate priors are progressively updated during development through exposure to sensory input and experience, becoming empirical priors via perceptual inference. Through Bayesian inference, prior probabilities are continuously revised in light of new sensory information, generating updated posterior probabilities. These posteriors, in turn, serve as priors for subsequent inferences, thereby supporting an ongoing cycle of learning and adaptation. This iterative process—where priors are successively refined into posteriors—facilitates the continuous optimization of the brain's internal model of the environment. During development, the child's brain persistently adjusts its predictions, becoming increasingly efficient at interpreting and anticipating sensory input on the basis of accumulated experience. This predictive-coding account of neurocognitive development is supported by both behavioral (Gopnik and Wellman, 2012; Köster et al., 2020) and neural evidence (Emberson et al., 2015; Jaffe-Dax et al., 2020; Zhang and Emberson, 2020) although further work is needed. For example, electrophysiological measures such as the mismatch negativity (MMN) and mismatch field (MMF) are expected to increase with age, reflecting reduced prediction error as internal models become more refined (Rapaport and Sowman, 2024). However, this pattern is not yet consistently observed across studies, age ranges, or experimental protocols (Gomot et al., 2000). More broadly, little is currently known about how children within the critical 10-year-old developmental window internally represent emotional prosody, or how changes in these internal representations relate to task performance.

A common experimental paradigm used to infer internal templates experimentally from stimulus-response data is psychophysical reverse correlation (Jack and Schyns, 2017). In a typical reverse correlation task, participants are presented with large quantities of stimuli whose physical characteristics are randomly manipulated using signal processing algorithms. Participants' responses to such random stimuli are used to "reverse-engineer" the internal representation they rely on during perceptual processing. In adult, reverse correlation paradigms have been used to uncover internal representations associated with facial expressions of emotions (Mangini and Biederman, 2004; Yan et al., 2023), as well as speech prosody (Goupil et al., 2021; Ponsot et al., 2018a).

While in principle, reverse correlation could be used with children to document the developmental trajectory of internal representations of emotional prosody, only a handful of studies have attempted to do so. For instance, Wang et al. (2023) have investigated internal representations of pitch contour in speech, complex tone and melody in children and adolescents aged 4 to 15 years. However, their analysis treated all participants as a single group, without examining developmental changes across age or including an adult control group. A major challenge in applying reverse correlation to child populations lies in the

typical procedural demands of such paradigms, which often require long experimental sessions with large numbers of trials (1000 trials in Wang et al., 2023), a design that is impractical for young or easily fatigued participants (see Adl Zarrabi et al., 2024 for a similar issue in brain-stroke patients). In Ponsot, Burred et al. (2019) developed a reverse-correlation paradigm specifically aimed at uncovering the representation of vocal smile, employing a relatively short two-alternative forced-choice (2AFC) task consisting of 600 trials. Although this paradigm was tested exclusively with adults, the robustness of its results suggested that reliable estimation could be achieved with as few as 150–300 trials (Merchie et al., 2024), making it a promising approach for use with younger populations.

1.3. Hypotheses and objectives

The overarching aim of this study is to investigate the internal representation of vocal smile in school-aged children (8–12 years).

Given the ongoing anatomical and functional maturation of the auditory cortex, particularly the development of sensitivity to acoustic features, we hypothesize that the internal representation of vocal smile is still under construction in this age group.

However, an alternative (and not mutually exclusive) possibility is that the auditory representation itself may be mature, but its use in perceptual tasks remains sub-optimal due to immature decision-making processes. This could manifest as reduced robustness in children's internal models, stemming from higher internal noise or response variability.

To disentangle these possibilities, we adapted an existing reverse-correlation task (Ponsot et al., 2018a) with a reduced number of trials while preserving task validity. This approach allows us to extract the following key measures:

- The **acoustic signature** of the internal representation of vocal smile, reflected in the frequency filters applied to the original neutral sound and its derived measure,
- The **typicality** of each child's model, quantified by its similarity to an adult reference template.
- The **robustness** of the internal representation, estimated through internal noise measures derived from responses consistency and bias.

2. Materials and methods

2.1. Population

46 children (25 females) aged from 8 to 12 (10.5 ± 1.5 years) and 42 adults (22 females); aged from 20 to 58 years (25.2 ± 8.9 years) participated in the study. All participants were native French speakers. To be included in this study, participants should not present (or have history of) neurological or psychiatric disorder, developmental delay, visual or auditory disorders, language/speech disorders or be under medication at the time of the study. Each participant, or their parents in the case of children, signed an informed consent form, and the protocol was approved by the Ethics Committee (PROSCEA2017/23; ID RCB: 2017- A00756–47). Two children and one adult were not included in the analyses because of missing data. Verbal and non-verbal developmental age were estimated using 4 Weschler Intelligence Scale for Children V subtests: matrix-reasoning, cubes, vocabulary, similarities for children (Wechsler, 2014).

2.2. Stimuli and experimental design

2.2.1. . Stimuli

Following the procedure of (Ponsot et al., 2018a), we recorded a single utterance of the vowel /a/ pronounced with a constant pitch (~ 122 Hz) by a male speaker. From that sound, a 500 ms stationary part was selected to create a stimulus with constant spectral energy. To generate reverse-correlation stimuli, we then randomized the spectral content of the baseline stimulus using spectral-domain equalizers consisting of a bank of 25 band-pass filters with fixed cutting frequencies and random gains. The filters' cutting frequencies were logarithmically spaced at frequency points ranging from 100 to 10,000 Hz. The filters' gain values (in dB) were drawn from Gaussian distributions (SD = 5 dB, clipped at ±2.5 SD). The procedure was implemented using the open-source CLEESE toolbox (Burred et al., 2019).

2.2.2. Procedure

Participants were presented with pairs of modified voices and, for each pair, were asked to choose the one produced with the greatest smile (2-interval, 1-alternative forced choice). In accordance with Merchie et al., 2024, adults were presented with 250 of such pairs (consisting of all-different random 500 stimuli). To reduce experimental session time for children, a convergence analysis was performed on adults' results (see supplementary materials, part I). This analysis showed that representations extracted after 150 trials showed sufficient similarity with those extracted with 250 trials, and so trial number was reduced to 150 (300 stimuli) for children. In addition to these 150 (children) or 250 (adults) trials, an additional 50 trials were repeated identically at the end of the experiment, a so-called "double-pass" procedure used as a way to estimate internal noise with the proxy of response consistency (Neri, 2010). Trials were presented in blocks of 50 pairs, resulting in 4 blocks for children (3 initial blocks + one repeated) and 6 blocks for adults (5 initial blocks + one repeated). Resulting in the last two blocks being rigorously the same, with the same trials in the same order.

Sounds were delivered using closed headphones (Beyerdynamics DT770) presented the stimuli dichotically (same signal in both ears) at an identical comfortable sound level (~ 70 dB SPL) to all children and adults.

2.2.3. Data analysis

Three key measures can be extracted from this task, the internal representation of vocal smile (acoustical features of the smiling filter), the typicality of children's internal representation (a single feature of the representation in comparison to adults' one) and internal noise (estimated from consistency and response bias).

2.2.3.1. Internal representation of vocal smile. Internal representation of vocal smile was estimated using the classification image technique (Murray, 2011). The individual representation consists of a frequency filter (a 25 points-vector), of the same shape as each of the reverse-correlation stimulus. This vector was calculated for each participant, as the mean filter (i.e., acoustic characteristics) of voices chosen as "smiling", from which the mean filter of the non-chosen voices was subtracted. This representation provides a gain for each of the 25 modulated frequencies of the sound, which describe in what spectral direction a sound should be modified in order to increase its probability to be chosen as smiling by the participant. If a specific gain is equal to zero, this means that no modulation of the sound at that specific frequency provides any cue for judging vocal smiles. To compare representations across participants, we normalized each individual representation by dividing it by its root mean square (RMS) value.

2.2.3.2. Typicality of internal representation. To synthetize the vocal smile representation into one single feature, the typicality of each child model was determined in comparison to adults' vocal smile model.

This measure was based on acoustic filter of vocal smile i.e., the gain associated with the 25 frequencies of the sound. First, the average acoustic filter of adults was computed by averaging at each of the 25 frequencies the gain value of each adult, resulting in a single representation, (25 gains values associated to the 25 frequencies of the sound). Then the absolute-value distance between child and adults' mean gains at each 25 frequency points was computed resulting in a distance for

each frequency. Then these distances were added together to obtain a single value ($distance_{child}$, Eq. (1)) representing the average distance between a child's representation and the average adult model (Eq. (1)).

The distance was then rescaled with respect to the children group to obtain a value of typicality between 0 and 1 (Eq. (2)), with typicality values approaching 0 indicating that the child's representation is far from adults, and typicality approaching 1 indicating their representation is very similar.

$$distance_{child} = \sum_{f=100}^{10000} \left| Gain_{f,adults} - Gain_{f,child} \right| \quad (1)$$

$$Typicality_{child} = 1 - \frac{(distance_{child} - min(distance_{children}))}{max(distance_{children}) - min(distance_{children})} \quad (2)$$

2.2.3.3. Internal noise estimation. Internal noise was estimated for each participant thanks to the double-pass procedure (Neri, 2010). First, two indices were computed from the comparison between two identical blocks (1) the probability pagree of choosing the same sound as most smiling between the two blocks (i.e., response consistency) and (2) probability pfirst of choosing the first sound over the second (i.e., response bias). Internal noise was then estimated from these two quantities using the simulation procedure of Adl Zarrabi et al., 2024. We simulated ideal observers with an arbitrary representation (ex. a simple scalar) and a grid of (true, known) internal noise σ_n (between 0 and 5) and bias b (between −5 and 5) values and let them encounter a single large (n = 10,000 trials) experiment, repeated 100 times. We then computed the average value of pagree and pfirst over these realizations for each value of σ_n and b, and store this in a lookup table. Given the empirical observation of probabilities pagree and pfirst for a real observer, the procedure can then find the pair of internal noise and bias that correspond to the closest pair of probabilities in the table.

Both analyses (internal representations and internal noise) were conducted using the open-source toolbox PALIN (https://github.com/neuro-team-femto/palin).

2.2.4. Statistical analysis

Statistical analyses were conducted with Rstudio 4.2.2 (R Core Team, 2021).

To compare adults and children's representations of vocal smile, two-tailed *t*-tests were performed between gains associated with the 25 frequencies of each group model, with Bonferroni correction. The same analysis was performed within each group in comparison to 0 to assess internal representation of children and adults.

To compare the formant frequencies (F1, F2, F3 and F4) in the vocal smile model of children and adults, we applied each of the participant's model (i.e., the acoustical filter corresponding to gain at each of the 25 frequencies of the sound) to the baseline sound, extracted the formants of the resulting sound (i.e., modified sound corresponding to the internal representation) using Praat (Boersma, 2002), and compared to those of the original sound using one-sample *t*-tests.

Internal representation typicality was computed for each child, and a Pearson correlation analysis was performed between Age and Typicality. Pearson's correlation analysis was computed between Typicality and EQ.

Internal Noise was computed for each child. Children were separated into two groups regarding their IN value, IN+ for IN > 3 (n = 14 children) and IN- for IN < 3 (n = 30 children).

A specific index was computed to evaluate F1 increase or decrease in each participant. First, for each child ΔF1 was computed, corresponding to the difference between each child F1 frequency (Hz) and F1 frequency of the original sound (555 Hz): $\Delta F1 = F1_{child} - 555$. Then to compare this value to the adults' one, (corresponding to the 'target' F1 modulation), ΔF1 was normalized regarding adults group using the min/max method: $norm(\Delta F1) = (\Delta F1_{child} - mean(\Delta F1_{adults}))/((max(\Delta F1_{adults}) - (min(\Delta F1_{adults})))$

Thus, norm(ΔF1) equal to 0 means that the child representation contains the same modulation of F1 frequency as adults' representation, a positive one a greater increase than adults, and a negative one a decrease compared to adults.

For each group (IN+ and IN-) two correlations were performed: the first one between norm(ΔF1) and Age (months) and between norm(ΔF1) and Typicality. A p-value correction was applied to correct for multiple comparisons. Spearman's correlations were made according to data distribution (normality was checked using Shapiro-wilk). Spearman's correlation were compared between groups using Zou's confidence intervals (CI) (R package: cocor) because of its acknowledgment of the asymmetry of sampling distributions for single correlations (Zou, 2007). Correlations were interpreted as significantly different if Zou's CI did not contain zero (Zou, 2007).

2.2.5. Data statement

Code and data for processing and statistical analysis are publicly available in an OSF project (https://osf.io/8b36h/)

3. Results

3.1. Internal representation of vocal smile in children

The averaged frequency filter underlying the evaluation of smile in the /a/ vowel is depicted in Fig. 1A. For both groups, the internal representation of vocal smile is characterized by a diminution of energy at low frequencies and an amplification of formants F1 and F2 (Fig. 1A and B).

To compare adults and children's representation of vocal smile, two-tailed *t*-tests were performed between the gains associated with the 25 frequencies of sounds for each group, with Bonferroni correction. No significant difference was found (see supplementary material, part II). In short, children appeared to be using the same auditory representation as adults to evaluate smiling voices.

The first four formants of adults and children's acoustic representation of vocal smile were compared to original sounds formants (see Table 1).

The smiling representation was mainly characterized by an increased frequency of formant F1 (Fig. 1A) ($t(43) = 14.03$, $p < .001$, $d = 2.11$) (Table 1). In the following sections, each participant's F1 modulation will be referred to as ΔF1. This corresponds to the difference between the participant's F1 value in the smiling filter condition and the original sound F1 (555 Hz).

3.2. Representation typicality

Typicality is a single feature to compare representations of vocal smile within children compared to adults (i.e., where adults are considered as the norm). As a child's typicality approaches 1, their internal representation becomes increasingly similar to those observed in adults.

Even if, as shown in Fig. 1, children display the same representation of vocal smile as adults, it remains heterogenous within the group (Fig. 2).

A Pearson' correlation analysis was performed between typicality and age (in months) in children. This analysis revealed a moderate positive correlation between the two variables ($t(42) = 3.1$, $p < .005$, $R = .43$), with an increase of typicality with children age (Fig. 3).

Typicality was also computed on specific segments of frequencies to investigate whether some frequency band would have different developmental dynamics. Results did not demonstrate difference and are detailed in supplementary material (part III).

Typicality was not correlated with EQ ($t(39) = .43$, $p = .67$, $R = .07$).

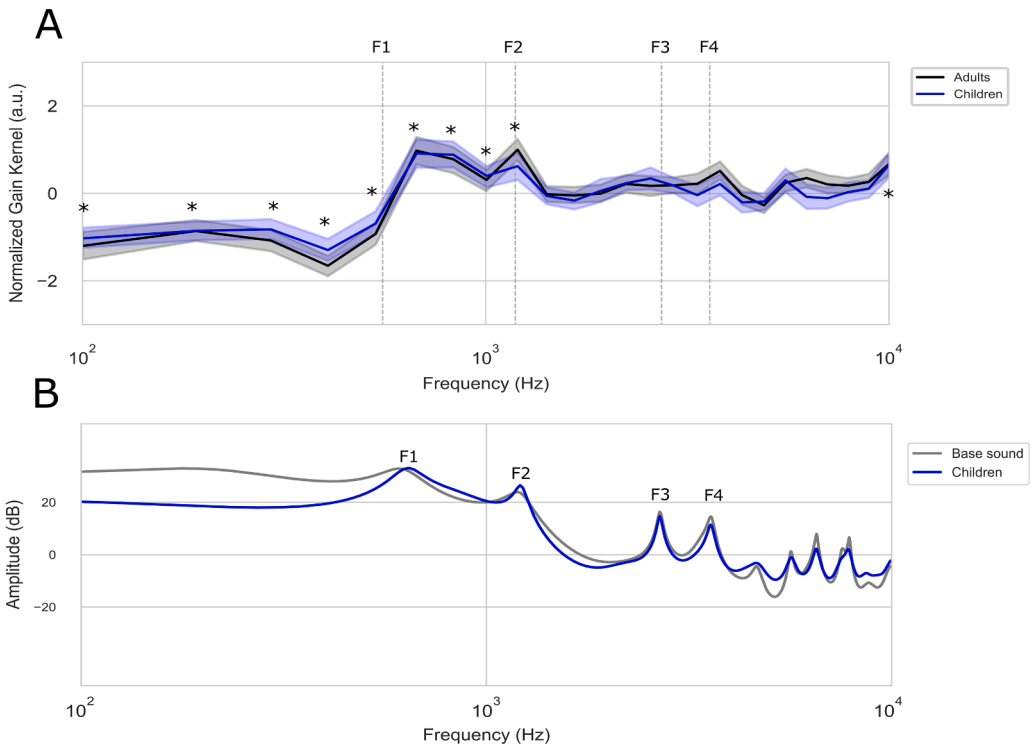

Fig. 1. A: Averaged filter underlying evaluation of vocal smile, as derived with reverse-correlation. Children are represented in blue and adults in black. A gain of 0 corresponds to the base sound and the curves represent the modification of the base sound to describe a vocal smile. Asterisks indicate significant differences from 0 (two-tailed; paired-sample *t*-tests, $p_{corrected} < 0.05$) (no difference between groups). B: Smiling Filter applied (here with a gain of 0.5) to the original voice, revealing the internal representation of vocal smile in children (in blue) compared to the base sound (in grey).

Table 1
Comparison of formants frequencies between the original sound and representation of vocal smile in children and in adults. Formants frequencies are presented in Hertz (mean ± sd). One-sample *t*-test were performed between formants of the original sound and the representation of vocal smile in children and in adults.

Original sound	**F1** (555 Hz)	**F2** (1183 Hz)	**F3** (2727 Hz)	**F4** (3591 Hz)
Children (n = 44)	644 ± 42 t(43) = 14.03, p < .001, d = 2.11	1192 ± 131 t(43) = .45, p > .1, d = .07	2712 ± 102 t(43) = −1.00, p > .1, d = −.15	3587 ± 88 t(43) = -.30, p > .1, d = −.05
Adults (n = 42)	638 ± 48 t(41) = 11.13, p < .001, d = 1.72	1185 ± 33 t(41) = .31, p > .1, d = .05	2735 ± 41 t(41) = 1.26, p > .1, d = .19	3596 ± 32 t(41) = 1.09, p > .1, d = .17

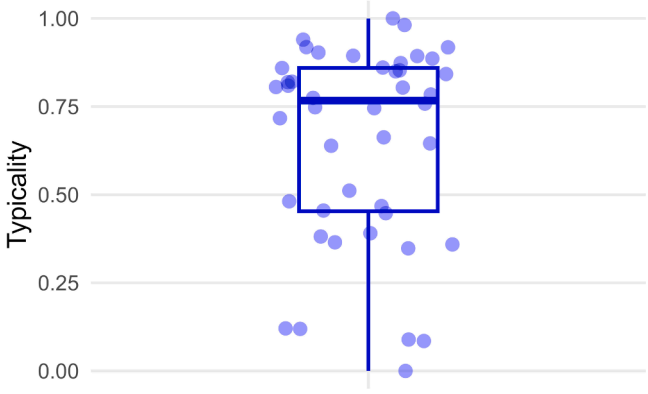

Fig. 2. Children's Typicality of vocal smile in comparison to adults.

3.3. *Representation robustness: Internal noise analysis*

Internal noise (IN) was estimated using the double-pass procedure Neri (2010), following the same methodology as Adl Zarrabi et al. (2024).

Due to a IN bimodal distribution (Fig. 4), children were separated into two groups: IN+ (IN > 3; n = 14 children) and IN- (IN < 3; n = 30 children). The threshold to separate both groups was chosen arbitrary regarding the bimodal distribution of IN data. The two groups do not differ in age (W = 162, p = .23), EQ (W = 154, p = .34), verbal (W = 179, p = .62) and nonverbal (W = 201, p = .25) developmental age. The IN+ children have a statistically greater IN than the IN- children (W = 420, p < .0001), and the IN- children mean IN did not statistically differ from adults (W = 559, p = .52).

To characterize the use of F1 modulation across development in children, a Spearman's correlation analysis was performed between norm(ΔF1) and Age (months) separately in each group (IN+ and IN-) (Fig. 5). A significant positive correlation was found (S = 176, p_{corr} = .045, R = .61) in the IN+ group, corresponding to an increase of F1 with age. No correlation (S = 4816, p_{corr} = 1.415, R = - .07) was found for the IN- group. These correlations were also significantly different (*Zou's CI*

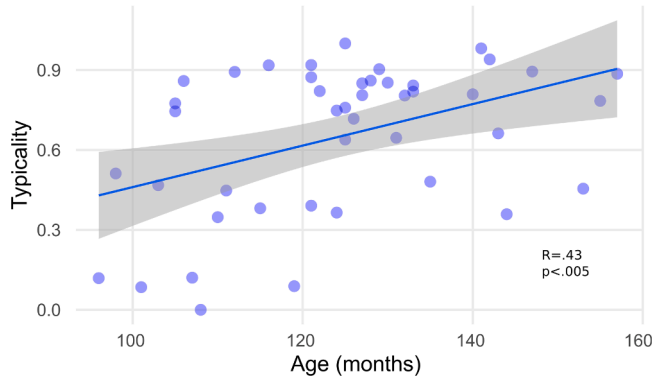

Fig. 3. Correlation between Typicality and Age (in months) in children. Typicality corresponds to the normalized distance between a child representation of vocal smile and adults' one.

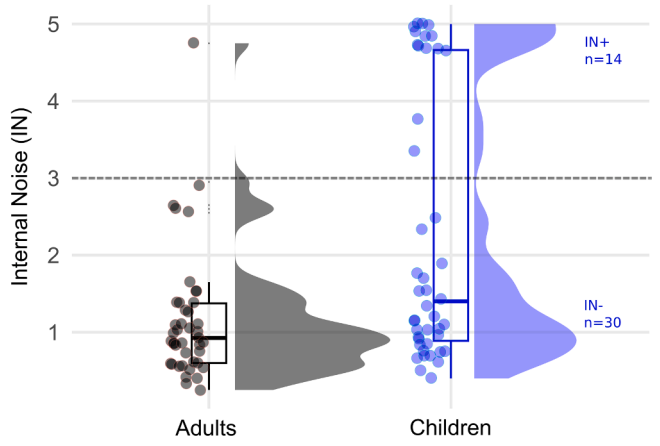

Fig. 4. Internal Noise (IN) bimodal distribution across children and adults.

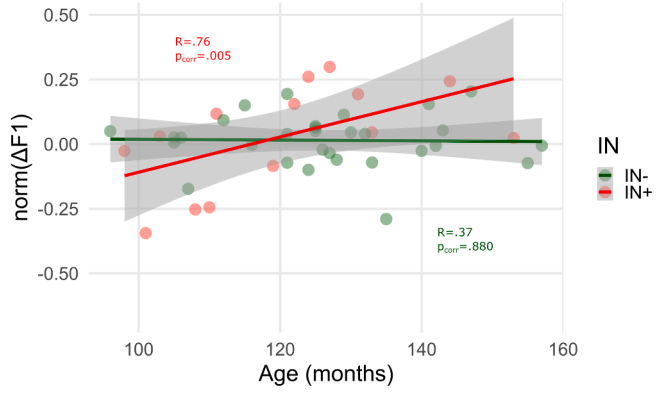

Fig. 5. Correlation between norm(ΔF1) and Age for IN+ and IN- children. Correlation is displayed in green for the IN- group and in red for the IN+ group. norm(ΔF1) corresponds to the normalized ΔF1 in comparison to adults (where ΔF1 is computed as the difference between F1 frequency of each child model and F1 of the original sound).

[0.07,1.1]).

A Spearman's correlation analysis was also performed between norm (ΔF1) and Typicality independently for each internal noise level group (IN+ and IN-) (Fig. 6). A significant positive correlation was found (S = 110, p_{corr} = .005, R = .76) for the IN+ group, while no correlation (S = 2824, p_{corr} = .880, R = .37) was found for the IN- group. For IN+

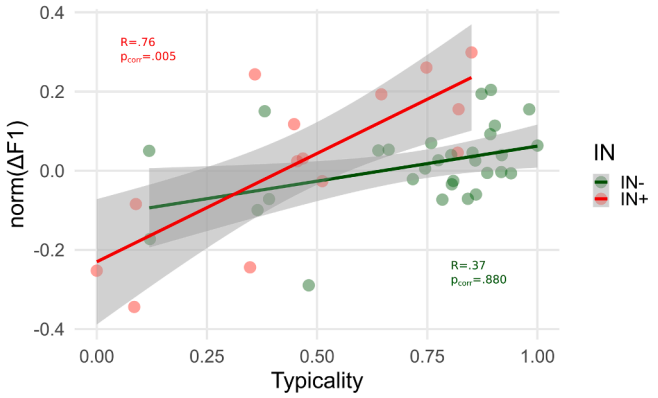

Fig. 6. Correlation analysis between Typicality and norm(ΔF1) for IN+ and IN- children. Correlation is displayed in green for the IN- group and in red for the IN+ group. norm(ΔF1) corresponds to the normalized ΔF1 in comparison to adults (where ΔF1 is computed as the difference between F1 frequency of each child model and F1 of the original sound). Typicality corresponds to the normalized distance between a child representation of vocal smile and adults' representation.

children, typicality was associated with an increase in F1 frequency, while, IN− children, typicality variation is not significantly correlated to F1 modulation. These correlations were not significantly different (*Zou's CI* [−0.08,.78]).

4. Discussion

The present study is, to our knowledge, the first to examine the development of the mental representation of vocal emotions in children. For this purpose, a reverse correlation method was adapted to obtain emotional internal representation and an associated measure of representation robustness (internal noise) in 8–12 years old children. Our main finding is that internal representation of vocal smile becomes increasingly refined with age and that this developmental change reflects the sharpening of the underlying rather than a lack of representation robustness in young children. In other words, younger children appear to possess a rudimentary version of the adult template, but this representation gains precision with age. Importantly, our results also identified a subgroup of children who exhibited a distinct developmental trajectory in acquiring the internal representation of the vocal smile.

Regarding children's internal representation of vocal smile, results replicate those found in adults (Merchie et al., 2024; Ponsot et al., 2018a), indicating a fine-tuned abilities to infer a primarily visual emotion (the smile) from acoustic cues in the voice. Small differences from Ponsot et al., 2018b study might be the consequence of the reduction in the number of trials in our experiment and the variability of participants. The mouth opening associated with smiling (Ekman and V. Friesen, 1978) modifies the emitted sound, leading to an increase of F1 frequency of the smiling voice. In line with prior findings (Frick, 1985; Ohala, 2009; El Haddad et al., 2015) the acoustic profile of the smiling voice in our study was mainly characterized by a reduction in low-frequency energy, increase energy of F1 and F2 and an increase in the first formant frequency, confirming that children's internal representations capture these characteristic spectral modulations. However, our results are specific to /a/ vowel while others studies were interested in a wide range of vowels (El Haddad et al., 2015; Podesva et al., 2015) and thus found modifications of others formants frequencies such as F2. As in adults, children's internal representation can thus be operationalized as a specific acoustic smiling filter.

Although children as a group exhibit the same general template as adults, individual differences emerged, consistent with a positive correlation between typicality and age. As depicted by the predictive

coding framework, internal representation of vocal smile becomes increasingly refined with experience (Rapaport and Sowman, 2024). Thus, the representation of that sensory input is not yet fully implemented, which may hinder the recognition of associated emotional prosody, as depicted by the literature. While basic vocal emotion recognition abilities emerge as early as 6 to 9 years of age (Matsumoto and Kishimoto, 1983), the point at which these skills reach adult-like proficiency remains debated (see review by Morningstar et al. (2018)). Behavioral data indicates that individuals aged 11 to 15 still perform below adult levels in vocal emotion recognition (Brosgole and Weisman, 1995; Chronaki et al., 2018), suggesting that these abilities continue to develop throughout adolescence and into adulthood. Furthermore, throughout this critical period, neuronal communication becomes more efficient due to an extended phase of synaptic pruning and myelination (Blakemore and Choudhury, 2006; Fischer and Bidell, 2007; Santos and Noggle, 2011). Notably, the prefrontal cortex undergoes prolonged development (Giedd et al., 1999; Huttenlocher and Dabholkar, 1997). This region plays a crucial role in identifying complex statistical patterns within sensory input and is also implicated in emotion recognition (ER) processes (Jaffe-Dax et al., 2020; Basirat et al., 2014). Specifically, neural responses in the dorsomedial prefrontal cortex and inferior frontal gyrus increase with age during vocal ER tasks, as observed in individuals aged 8 to 19 years (Morningstar et al., 2019). This prolonged maturation may contribute to the refinement of predictive brain processes throughout development (Krogh and H.Johnson, 2013).

Internal representation becomes more precise with age. However, this improvement may also reflect greater task engagement in older children. In this case, the increase of accuracy would not necessarily rely on specific changes of the internal representation of vocal smile itself, but rather on a reduction of internal noise. Indeed, growing up, children may simply have better access to their internal representation by reducing internal noise (see Adl Zarrabi et al. (2024) for insights on representation + noise model). Indeed, internal noise governs behavioral variability in how children apply their template of vocal smile in perceptual task, conversely to other psychophysical studies where noise refers to intrinsic noise of the stimulus (Jack and Schyns, 2017). Consistent with this interpretation, the IN+ and IN− groups did not differ in age, suggesting that children who perform poorly (i.e., exhibited higher internal noise) are not simply the youngest participants We therefore divided participants into two subgroups IN+ and IN− to investigate whether differences in acoustic cue utilization could explain performance variability.

Correlations between ΔF1 and Age revealed distinct developmental patterns between groups. IN- children have already reached adult-like F1 modulation patterns, demonstrating that even 8-year-olds can use F1 variation as a diagnostic cue in the perception of vocal smile. This aligns with findings by Keough et al. (2015) who observed increased F1 values during smile production for the /a/ vowel. Meanwhile, as IN+ children grow up, they begin to use F1 augmentation as a cue to perceive vocal smiles. Younger IN+ children tend not to rely on F1 augmentation for this characterization, whereas older children use it even beyond the template observed in adults. These results are showing that IN+ children are still improving their ability to extract accurate acoustic cues from emotional prosody sensory inputs, even until overpassing the F1 modulation of adults' template. A data-driven approach combining speech phonetics, acoustic features and electrocochleography (ECOG) recordings revealed how phonetic representations are encoded in the human Superior Temporal Gyrus (STG) (Mesgarani et al., 2014). F1 and F2 variability occurs within the STG where single sites respond to integrated formant patterns. Continued maturation of the STG during late childhood could therefore account for the distinct developmental dynamics observed in the IN+ subgroup. Future studies could test this hypothesis using auditory ERP markers of STG maturation, as significant changes are known to occur between 8 and 12 years.

A second analysis explored the relationship between Typicality and F1 modulation. Since IN- children have generally already achieved F1 increase as a cue for vocal smile, yet continue to refine their internal representations, they may rely on more subtle acoustic indices to further enhance their perceptual accuracy. Regarding F1 modulation, IN- children achieved a plateau equivalent to adults' F1 modulation, thus typicality is no longer modulated by F1 increase. Indeed, there is no correlation between Typicality and ΔF1 for IN- group. We expect IN- children, associated with global neural maturation (Giedd et al., 1999), to rely on frequencies modulation energy or others accurate acoustic cues to categorize a voice as smiling such as more subtle variations on other frequencies that are not caught by our measures. Meanwhile, for IN+ children, Typicality is strongly associated with ΔF1, these children may focus on the most salient acoustic cue to categorize a voice as smiling, which is F1 modulation. We could hypothesize that the IN+ group could still improve its internal model of vocal smile, even if no link has been made between IN and verbal or nonverbal developmental age. As the IN- children have the same amount of IN than adults, we could expect the IN+ children to reduce this variability in responses by growing older. This group would follow another or delayed developmental pathway. Especially in the scope of emotional prosody recognition, this developmental pathway could be the consequence of heterogeneity within population. A recent study on individual differences in multimodal emotion recognition abilities suggests that variability in this domain may manifest at multiple stages of decisional processing including the extraction (middle Superior Temporal Gyrus), integration (posterior Superior Temporal Sulcus), and evaluation (Inferior Frontal Cortex) of emotional information (Laukka et al., 2024). These decisional processing stages could be linked to internal noise measure since it's estimated from consistency and response bias. Indeed, as proposed by Vilidaite et al. (2017) the measured internal noise (thanks to the double pass procedure) may be a late decision-making noise that influences behavior at the level of execute function. Regarding predictive coding hypothesis on prediction error diminution, children might not all have the same innate priors and certainly not the same sensory inputs that enrich their inner representation of a specific vocal emotion.

Several limitations should be acknowledged. This study was a part of a larger protocol, and fatigability may have affected performance in some children. Group sizes were unequal (IN−: $n = 30$; IN+: $n = 14$), which limits statistical power for between-group comparisons. Additionally, representation robustness was estimated from a limited number of late-trial responses, which may underestimate stability. Despite these constraints, this adapted reverse-correlation method is thought to be promising and effective to evaluate the internal sensory representation of speech emotional prosody. This technique could thus be implemented with many others vocal emotions since it has been demonstrated that not all basic emotions achieve adult-like accuracy scores at the same time (Riddell et al., 2024). Taken together, our findings provide intriguing possibilities for exploring the nuances of perception in prosody within neurodevelopmental disorders such as Autism Spectrum Disorders (ASD). As described by Keating et al. (2023) for visual emotions in ASD adults, atypicalities described in the perception of emotions in ASD could be linked to a different utilization of these representations instead of a specific deficit in emotional processing. Longitudinal assessment of these mechanisms during development could therefore provide critical insight into the construction of emotional perception networks in ASD.

5. Conclusion

To our knowledge, this study is the first to implement a straightforward yet effective method for assessing vocal emotional representations and associated internal noise in school-aged children. Based on a sample of 44 children aged 8 to 12 years, our findings reveal that with increasing age, children progressively refine their internal representation of vocal smile, gradually converging toward the adult perceptual template. While younger children exhibit less precise representations, a subset appears to follow an alternative developmental trajectory,

characterized by decreased representation robustness interfering with representation accuracy. These results highlight the heterogeneity of emotional voice processing during development and open new avenues for identifying early markers of atypical socio-emotional perception.

Ethics statement

The protocol received approval from local (CER-TP 2023-09-02) and national (PROSCEA2017/23; ID RCB: 2017-A00756-47) Ethics Committees. Each participant or their parents signed an informed consent form.

Funding

This work was supported by the INSERM (National Institute for Health and Medical Research), the ANR (ANR-19-CE37-0022-01), the Region CVL. The funding sources were not involved in the study design, data collection and analysis, or the writing of the report.

CRediT authorship contribution statement

Zoé Ranty: Writing – review & editing, Writing – original draft, Visualization, Software, Methodology, Investigation, Formal analysis, Data curation. **Annabelle Merchie:** Writing – review & editing, Software, Methodology, Data curation, Conceptualization. **Jean-Julien Aucouturier:** Writing – review & editing, Validation, Software, Methodology, Formal analysis, Conceptualization. **Marie Gomot:** Writing – review & editing, Writing – original draft, Validation, Supervision, Resources, Project administration, Methodology, Funding acquisition, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

We thank all the children and their parents for their time and participation in this study. We would like to thank Camille Bataillon for her help in children recruitment and data acquisition and Aynaz Adl Zarrabi for her insights on methodological aspects.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.specom.2026.103452.

Data availability

Data will be made available on request.

References

Adl Zarrabi, A., Jeulin, M., Bardet, P., Commère, P., Naccache, L., Aucouturier, J.-J., Ponsot, E., Villain, M., 2024. A simple psychophysical procedure separates representational and noise components in impairments of speech prosody perception after right-hemisphere stroke. Sci. Rep. 14 (1), 15194. https://doi.org/10.1038/s41598-024-64295-y.

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), e83657. https://doi.org/10.1371/journal.pone.0083657.

Amorim, M., Anikin, A., Mendes, A.J., Lima, C.F., Kotz, S.A., Pinheiro, A.P., 2021. Changes in vocal emotion recognition across the life span. Emot. 21 (2), 315–325. https://doi.org/10.1037/emo0000692.

Badcock, P.B., Friston, K.J., Ramstead, M.J.D., Ploeger, A., Hohwy, J., 2019. The hierarchically mechanistic mind : an evolutionary systems theory of the human brain, cognition, and behavior. Cogn. Affect. Behav. Neurosci. 19 (6), 1319–1351. https://doi.org/10.3758/s13415-019-00721-3.

Barthel, H, Quené, H. Acoustic-phonetic properties of smiling revised : measurements on a natural video corpus. https://www.academia.edu/14939441/Acoustic-phonetic_properties_of_smiling_revised_-_measurements_on_a_natural_video_corpus.

Basirat, A., Dehaene, S., Dehaene-Lambertz, G., 2014. A hierarchy of cortical responses to sequence violations in three-month-old infants. Cognition 132 (2), 137–150. https://doi.org/10.1016/j.cognition.2014.03.013.

Basso, F., Oullier, O., 2010. Smile down the phone" : extending the effects of smiles to vocal social interactions. Behav. Brain Sci. 33 (6), 435–436. https://doi.org/10.1017/S0140525X10001469.

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. https://doi.org/10.1111/j.1469-7610.2006.01611.x.

Boersma, P., 2002. Praat, a system for doing phonetics by computer. Glot Int. 5, 341–345.

Brosgole, L., Weisman, J., 1995. Mood recognition across the ages. Int. J. Neurosci. 82, 169–189. https://doi.org/10.3109/00207459508999800.

Bryant, G.A., Barrett, H.C., 2007. Recognizing intentions in infant-directed speech : evidence for universals. Psychol. Sci. 18 (8), 746–751. https://doi.org/10.1111/j.1467-9280.2007.01970.x.

Burred, J.J., Ponsot, E., Goupil, L., Liuni, M., Aucouturier, J.-J., 2019. CLEESE: an open-source audio-transformation toolbox for data-driven experiments in speech and music cognition. PLOS One 14 (4), e0205943. https://doi.org/10.1371/journal.pone.0205943.

Chronaki, G., Wigelsworth, M., Pell, M.D., Kotz, S.A., 2018. The development of cross-cultural recognition of vocal emotion during childhood and adolescence. Sci. Rep. 8 (1), 8659. https://doi.org/10.1038/s41598-018-26889-1.

Eggermont, J.J., Moore, J.K., 2012. Morphological and Functional Development of the Auditory Nervous System. In: Werner, L., Fay, R.R., Popper (Éds.), A.N. (Eds.), *Human Auditory Development*. Springer, pp. 61–105. https://doi.org/10.1007/978-1-4614-1421-6_3.

Eibl-Eibesfeldt, I., 2017. Human Ethology. Routledge. https://doi.org/10.4324/9780203789544.

Ekman, P., 1992. Facial expressions of emotion : an old controversy and new findings. Philos. Trans. R. Soc. Lond. B Biol. Sci. 335 (1273), 63–69. https://doi.org/10.1098/rstb.1992.0008.

Ekman, P., & V. Friesen, W. (1978). Facial Action Coding System. https://books.google.fr/books/about/Facial_Action_Coding_System.html?hl=fr&id=08l6wgEACAAJ&redir_esc=y Consulting Psychologists Press.

El Haddad, K., Dupont, S., d'Alessandro, N., Dutoit, T., 2015. An HMM-based speech-smile synthesis system : an approach for amusement synthesis. In: Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops, 5, pp. 1–6. https://doi.org/10.1109/FG.2015.7284858.

Emberson, L.L., Richards, J.E., Aslin, R.N., 2015. Top-down modulation in the infant brain : learning-induced expectations rapidly affect the sensory cortex at 6 months. Proc. Natl. Acad. Sci. 112 (31), 9585–9590. https://doi.org/10.1073/pnas.1510343112.

Filippa, M., Lima, D., Grandjean, A., Labbé, C., Coll, S.Y., Gentaz, E., Grandjean, D.M., 2022. Emotional prosody recognition enhances and progressively complexifies from childhood to adolescence. Sci. Rep. 12 (1), 17144. https://doi.org/10.1038/s41598-022-21554-0.

Filippi, P., Ocklenburg, S., Bowling, D.L., Heege, L., Güntürkün, O., Newen, A., De Boer, B., 2017. More than words (and faces) : evidence for a Stroop effect of prosody in emotion word processing. Cogn. Emot. 31 (5), 879–891. https://doi.org/10.1080/02699931.2016.1177489.

Fischer, K.W., Bidell, T.R., 2007. Dynamic development of action and thought. Handbook of Child Psychology. John Wiley & Sons Ltd. https://doi.org/10.1002/9780470147658.chpsy0107.

Frick, R.W., 1985. Communicating emotion : the role of prosodic features. Psychol. Bull. 97 (3), 412–429. https://doi.org/10.1037/0033-2909.97.3.412.

Giedd, J.N., Blumenthal, J., Jeffries, N.O., Castellanos, F.X., Liu, H., Zijdenbos, A., Paus, T., Evans, A.C., Rapoport, J.L., 1999. Brain development during childhood and adolescence : a longitudinal MRI study. Nat. Neurosci. 2 (10), 861–863. https://doi.org/10.1038/13158.

Golan, O., Baron-Cohen, S., Hill, J.J., Rutherford, M.D., 2007. The 'reading the mind in the voice' Test-revised : a study of complex emotion recognition in adults with and without Autism spectrum conditions. J. Autism Dev. Disord. 37 (6), 1096–1106. https://doi.org/10.1007/s10803-006-0252-5.

Gomot, M., Giard, M.-H., Roux, S., Barthélémy, C., Bruneau, N., 2000. Maturation of frontal and temporal components of mismatch negativity (MMN) in children. NeuroReport 11 (14), 3109–3112.

Goodfellow, S., Nowicki Jr., S., 2009. Social adjustment, academic adjustment, and the ability to identify emotion in facial expressions of 7-year-old children. J. Genet. Psychol. 170 (3), 234–243. https://doi.org/10.1080/00221320903218281.

Gopnik, A., Wellman, H.M., 2012. Reconstructing constructivism : causal models, bayesian learning mechanisms, and the theory theory. Psychol. Bull. 138 (6), 1085–1108. https://doi.org/10.1037/a0028044.

Goupil, L., Ponsot, E., Richardson, D., Reyes, G., Aucouturier, J.J., 2021. Listeners' perceptions of the certainty and honesty of a speaker are associated with a common prosodic signature. Nat. Communicat. 12 (1), 861. https://doi.org/10.1038/s41467-020-20649-4.

Grossmann, T., Striano, T., Friederici, A.D., 2005. Infants' electric brain responses to emotional prosody. NeuroReport 16 (16), 1825–1828. https://doi.org/10.1097/01.wnr.0000185964.34336.b1.

Herba, C., Phillips, M., 2004. Annotation : development of facial expression recognition from childhood to adolescence: behavioural and neurological perspectives. J. Child Psychol. Psychiatry 45 (7), 1185–1198. https://doi.org/10.1111/j.1469-7610.2004.00316.x.

Hou, X., Zhang, P., Mo, L., Peng, C., Zhang, D., 2024. Sensitivity to vocal emotions emerges in newborns at 37 weeks gestational age. eLife 13, RP95393. https://doi.org/10.7554/eLife.95393.

Huttenlocher, P.R., Dabholkar, A.S., 1997. Regional differences in synaptogenesis in human cerebral cortex. J. Comp. Neurol. 387 (2), 167–178. https://doi.org/10.1002/(sici)1096-9861(19971020)387:2<167::aid−ne1>3.0.co;2-z.

Jack, R.E., Schyns, P.G., 2017. Toward a social psychophysics of face communication. Annu. Rev. Psychol. 68, 269–297. https://doi.org/10.1146/annurev-psych-010416-044242.

Jaffe-Dax, S., Boldin, A.M., Daw, N.D., Emberson, L.L., 2020. A computational role for top–Down modulation from frontal cortex in infancy. J. Cogn. Neurosci. 32 (3), 508–514. https://doi.org/10.1162/jocn_a_01497.

Keating, C.T., Ichijo, E., Cook, J.L., 2023. Autistic adults exhibit highly precise representations of others' emotions but a reduced influence of emotion representations on emotion recognition accuracy. Sci. Rep. 13 (1), 11875. https://doi.org/10.1038/s41598-023-39070-0.

Keough, M., Ozburn, A., McClay, E.K., Schwan, M.D., Schellenberg, M., Akinbo, S., Gick, B., 2015. Acoustic and articulatory qualities of smiled speech. Can. Acoust. 43 (3), 3. https://jcaa.caa-aca.ca/index.php/jcaa/article/view/2797.

Köster, M., Kayhan, E., Langeloh, M., Hoehl, S., 2020. Making sense of the world : infant learning from a predictive processing perspective. Perspect. Psychol. Sci. 15 (3), 562–571. https://doi.org/10.1177/1745691619895071.

Krogh, L.Vlach, H.Johnson, S.P, 2013. Statistical Learning Across Development : Flexible Yet Constrained. Front. Psychology. 3. https://doi.org/10.3389/fpsyg.2012.00598.

Laukka, P., Månsson, K.N.T., Cortes, D.S., Manzouri, A., Frick, A., Fredborg, W., Fischer, H., 2024. Neural correlates of individual differences in multimodal emotion recognition ability. Cortex 175, 1–11. https://doi.org/10.1016/j.cortex.2024.03.009.

Mangini, M.C., Biederman, I., 2004. Making the ineffable explicit : estimating the information employed for face classifications. Cogn. Sci. 28 (2), 209–226. https://doi.org/10.1207/s15516709cog2802_4.

Matsumoto, D., Kishimoto, H., 1983. Developmental characteristics in judgments of emotion from nonverbal vocal cues. Int. J. Intercult. Relat. 7 (4), 415–424. https://doi.org/10.1016/0147-1767(83)90047-0.

Matsumoto, D., Lee, M., 1993. Consciousness, volition, and the neuropsychology of facial expressions of emotion. Conscious Cogn. 2 (3), 237–254. https://doi.org/10.1006/ccog.1993.1022.

Menezes, C., Igarashi, Y., 2006 December. The speech laugh spectrum. Proceedings of the 7th International Seminar on Speech Production (ISSP), pp. 517–524.

Merchie, A., Ranty, Z., Aguillon-Hernandez, N., Aucouturier, J.-J., Wardak, C., Gomot, M., 2024. Emotional contagion to vocal smile revealed by combined pupil reactivity and motor resonance. Sci. Rep. 14 (1), 1–12. https://doi.org/10.1038/s41598-024-74848-w.

Mesgarani, N., Cheung, C., Johnson, K., Chang, E.F., 2014. Phonetic feature encoding in Human superior temporal gyrus. Science 343 (6174), 1006–1010. https://doi.org/10.1126/science.1245994.

Moore, J.K., Guan, Y.-L., 2001. Cytoarchitectural and axonal maturation in Human auditory cortex. J. Assoc. Res. Otolaryngol. 2 (4), 297–311. https://doi.org/10.1007/s101620010052.

Morningstar, M., Mattson, W.I., Venticinque, J., Singer, S., Selvaraj, B., Hu, H.H., Nelson, E.E., 2019. Age-related differences in neural activation and functional connectivity during the processing of vocal prosody in adolescence. Cogn. Affect. Behav. Neurosci. 19 (6), 1418–1432. https://doi.org/10.3758/s13415-019-00742-y.

Morningstar, M., Nelson, E.E., Dirks, M.A., 2018. Maturation of vocal emotion recognition : insights from the developmental and neuroimaging literature. Neurosci. Biobehav. Rev. 90, 221–230. https://doi.org/10.1016/j.neubiorev.2018.04.019.

Murray, R.F., 2011. Classification images : a review. J. Vis. 11 (5), 2. https://doi.org/10.1167/11.5.2.

Neri, P., 2010. How inherently noisy is human sensory processing? Psychon. Bull. Rev. 17 (6), 802–808. https://doi.org/10.3758/PBR.17.6.802.

Ohala, J.J., 2009. An Ethological Perspective on Common Cross-Language Utilization of F₀ of Voice. Phonet. 41 (1), 1–16. https://doi.org/10.1159/000261706.

Pegg, J.E., Werker, J.F., McLeod, P.J., 1992. Preference for infant-directed over adult-directed speech : evidence from 7-week-old infants. Infant Behav. Dev. 15 (3), 325–345. https://doi.org/10.1016/0163-6383(92)80003-D.

Pell, M., Monetta, L., Paulmann, S., Kotz, S., 2009. Recognizing emotions in a foreign language. J. Nonverbal. Behav. 33, 107–120. https://doi.org/10.1007/s10919-008-0065-7.

Podesva, R., Callier, P., Voigt, R., & Jurafsky, D. (2015, August). The connection between smiling and GOAT fronting: Embodied affect in sociophonetic variation. In *ICPhS*.

Ponsot, E., Arias, P., Aucouturier, J.-J., 2018a. Uncovering mental representations of smiled speech using reverse correlation. J. Acoust. Soc. Am. 143 (1), EL19–EL24. https://doi.org/10.1121/1.5020989.

Ponsot, E., Burred, J.J., Belin, P., Aucouturier, J.-J., 2018b. Cracking the social code of speech prosody using reverse correlation. Proc. Natl. Acad. Sci. 115 (15), 3972–3977. https://doi.org/10.1073/pnas.1716090115.

Ponton, C.W., Eggermont, J.J., Kwong, B., Don, M., 2000. Maturation of human central auditory system activity : evidence from multi-channel evoked potentials. Clin. Neurophysiol. 111 (2), 220–236. https://doi.org/10.1016/s1388-2457(99)00236-9.

Pujol, R., Lavigne-rebillard, M., Uziel, A., 1991. Development of the human cochlea. Acta Otolaryngol. 111 (sup482), 7–13. https://doi.org/10.3109/00016489109128023.

Quam, C., Swingley, D., 2012. Development in children's interpretation of pitch cues to emotions. Child Dev. 83 (1), 236–250. https://doi.org/10.1111/j.1467-8624.2011.01700.x.

R Core Team, 2021. R: A language and environment for statistical ## computing. R Foundation for Statistical Computing, Vienna, Austria. https://www.R-project.org/.

Rapaport, H., Sowman, P.F., 2024. Examining predictive coding accounts of typical and autistic neurocognitive development. Neurosci. Biobehav. Rev. 167, 105905. https://doi.org/10.1016/j.neubiorev.2024.105905.

Riddell, C., Nikolić, M., Dusseldorp, E., Kret, M.E., 2024. Age-related changes in emotion recognition across childhood : a meta-analytic review. Psychol. Bull. 150 (9), 1094–1117. https://doi.org/10.1037/bul0000442.

Rychlowska, M., Jack, R.E., Garrod, O.G.B., Schyns, P.G., Martin, J.D., Niedenthal, P.M., 2017. Functional smiles : tools for love, sympathy, and war. Psychol. Sci. 28 (9), 1259–1270. https://doi.org/10.1177/0956797617706082.

Santos, E., Noggle, C.A., 2011. Synaptic pruning. Éds.. In: Goldstein, S., Naglieri, J.A. (Eds.), Encyclopedia of Child Behavior and Development. Springer, US, pp. 1464–1465. https://doi.org/10.1007/978-0-387-79061-9_2856.

Sauter, D.A., Eisner, F., Ekman, P., Scott, S.K., 2010. Cross-cultural recognition of basic emotions through nonverbal emotional vocalizations. Proc. Natl. Acad. Sci. U.S.A. 107 (6), 2408–2412. https://doi.org/10.1073/pnas.0908239106.

Sauter, D.A., Panattoni, C., Happé, F., 2013. Children's recognition of emotions from vocal cues. Br. J. Dev. Psychol. 31 (1), 97–113. https://doi.org/10.1111/j.2044-835X.2012.02081.x.

Torres, M.M., Domitrovich, C.E., Bierman, K.L., 2015. Preschool interpersonal relationships predict kindergarten achievement : mediated by gains in emotion knowledge. J. Appl. Dev. Psychol. 39, 44–52. https://doi.org/10.1016/j.appdev.2015.04.008.

Vilidaite, G., Yu, M., Baker, D.H., 2017. Internal noise estimates correlate with autistic traits. Autism Res. 10 (8), 1384–1391. https://doi.org/10.1002/aur.1781.

Wang, L., Ong, J.H., Ponsot, E., Hou, Q., Jiang, C., Liu, F., 2023. Mental representations of speech and musical pitch contours reveal a diversity of profiles in autism spectrum disorder. Autism 27 (3), 629–646. https://doi.org/10.1177/13623613221111207.

Wechsler, D., 2014. Wechsler Intelligence Scale for Children–Fifth Edition (WISC-V) [Database record]. APA PsycTests. https://doi.org/10.1037/t79359-000.

Widen, S.C., 2020. Emotion perception and recognition. The Encyclopedia of Child and Adolescent Development. John Wiley & Sons Ltd., pp. 1–11. https://doi.org/10.1002/9781119171492.wecad171

Yan, S., Soladié, C., Aucouturier, J.-J., Seguier, R., 2023. Combining GAN with reverse correlation to construct personalized facial expressions. PLOS One 18 (8), e0290612. https://doi.org/10.1371/journal.pone.0290612.

Zhang, F., Emberson, L.L., 2020. Using pupillometry to investigate predictive processes in infancy. Infancy 25 (6), 758–780. https://doi.org/10.1111/infa.12358.

Zou, G.Y., 2007. Toward using confidence intervals to compare correlations. Psychol. Methods 12 (4), 399–413. https://doi.org/10.1037/1082-989X.12.4.399.