Emotion

The Cross-Modal Transfer of Emotional Information From Voices to Faces in 5-, 8- and 10-Year-Old Children and Adults: An Eye-Tracking Study

Amaya Palama, Jennifer Malsert, Didier Grandjean, David Sander, and Edouard Gentaz

Online First Publication, June 25, 2020. http://dx.doi.org/10.1037/emo0000758

CITATION
Palama, A., Malsert, J., Grandjean, D., Sander, D., & Gentaz, E. (2020, June 25). The Cross-Modal Transfer of Emotional Information From Voices to Faces in 5-, 8- and 10-Year-Old Children and Adults: An Eye-Tracking Study. *Emotion*. Advance online publication. http://dx.doi.org/10.1037/emo0000758

American Psychological Association

Emotion

© 2020 American Psychological Association
ISSN: 1528-3542

http://dx.doi.org/10.1037/emo0000758

The Cross-Modal Transfer of Emotional Information From Voices to Faces in 5-, 8- and 10-Year-Old Children and Adults: An Eye-Tracking Study

Amaya Palama, Jennifer Malsert, Didier Grandjean, and David Sander
University of Geneva

Edouard Gentaz
University of Geneva and The French National Centre for Scientific Research (CNRS), Paris, France

The present study examined the evolution of emotional cross-modal transfer throughout childhood compared to adulthood, using an experimental design first used with infants. We studied whether verbal children spontaneously look at emotional faces differently depending on the emotional voices previously heard, demonstrating a real intrinsic understanding of the emotion. Thus, sequences of emotional (happy vs. angry) cross-modal transfer were individually presented to 5-, 8- and 10-year-old children and adults. Spontaneous ocular behaviors toward the visual stimuli were recorded by eye-tracking. Results of the emotional cross-modal transfer suggested that participants looked spontaneously longer at the congruent face. However, this result was significantly revealed only as of age 8 with the happy voice and as of age 10 with the angry voice. Thus, the modulation of behavior indicators related to the control of the ability to extract amodal emotional information and spontaneously match the congruent information seems to increase with age and depends on the specific emotion presented.

Keywords: emotion, cross-modal transfer, development

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Emotional face, voice, and body expressions play a fundamental role in social communication. Emotions are important for interaction and make it possible to transmit one's internal state and intentions to others (Russell & Dols, 2017). Researchers have used many different instruments (tests, questionnaires, interviews, etc.) to measure the perception and understanding of emotions from early childhood to adulthood, and most require verbal abilities (for a review, see Castro, Cheng, Halberstadt, & Grühn, 2016). However, some populations, such as infants, some children with neurodevelopmental disorders, and non-native speakers, have some difficulty using language or receiving instructions. Thus, it is relevant to develop tasks that the participants can respond to independently of their level of receptive and expressive language. Consequently, some studies collect and analyze spontaneous oculomotor activity. In this case, researchers examine whether stimulus factors are associated with a specific oculomotor activity. Therefore, one way to study emotional understanding without using language skills would be to use the emotional successive cross-modal transfer task proposed to infants by Palama, Malsert, and Gentaz (2018). Indeed, the interest in studying emotional understanding is to limit the influence of language ability in cross-modal transfer. Cross-modal transfer does not require additional processes linked to language abilities, which may modify this operation and thus emotion understanding (Birch & Lefford, 1963; Bryant, 1974). In this study we want to assess emotional understanding using nonverbal instructions and responses in children and adults to explore the typical development of this ability.

The review of the literature on infancy reveals that the ability to recognize facial expressions emerges early, with infants' capacities to discriminate facial expressions at around 6–7 months of age (for a review, see Bayet, Pascalis, & Gentaz, 2014). Moreover, most intermodal matching studies show a preference for congruent stimuli from 3 to 7 months of age (Godard, Baudouin, Schaal, & Durand, 2016; Kahana-Kalman & Walker-Andrews, 2001; Montague & Walker-Andrews, 2002; Soken & Pick, 1992; Vaillant-Molina, Bahrick, & Flom, 2013; Walker, 1982; Zieber, Kangas, Hock, & Bhatt, 2014). However, the discrimination of two emotional faces by infants gives no information on their real understanding of the emotion per se; it merely shows they can perceptively differentiate stimuli like the visual characteristics of a face. Palama et al. (2018) overcame this difficulty by using a successive cross-modal transfer paradigm from emotional voices to emotional faces. Thus, the ability to transfer cross-modally from emotional voices to emotional faces (angry or happy) was examined in

Amaya Palama, Jennifer Malsert, Didier Grandjean, and David Sander, Faculty of Psychology and Sciences of Education and Swiss Center for Affective Sciences, University of Geneva; Edouard Gentaz, Faculty of Psychology and Sciences of Education and Swiss Center for Affective Sciences, University of Geneva, and The French National Centre for Scientific Research (CNRS), Paris, France.

The authors are grateful to the schools, teachers, children and their families, as well as the adult participants for their involvement and participation. We would also like to thank Margaux Pimont, Coralie Vouillon, and Kelly Canizares-Marin for their contribution, and finally Susan Campbell who provided English editing. This research was supported by the Swiss National Fund for research Grant 100019-156073 awarded to Edouard Gentaz. The authors had no personal or financial conflicts of interest.

Correspondence concerning this article should be addressed to Edouard Gentaz, Faculty of Psychology and Sciences of Education, University of Geneva, Uni Mail, 40 Boulevard du Pont-d'Arve, CH-1211, Genève 4, Switzerland. E-mail: edouard.gentaz@unige.ch

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

6-month-old infants. Results suggested no difference in infants' looking time at the happy or angry face after listening to the neutral voice or the angry voice, whereas, after listening to the happy voice, the incongruent angry face (the mouth area in particular) was watched longer than the congruent happy face. These results revealed that a cross-modal transfer (from the auditory to the visual modality) is possible as of 6 months of age after the presentation of a happy voice. Using the same paradigm in 2-, 4-, and 6-month-old infants, the results of Palama, Malsert, and Gentaz (2020) confirm the result found at 6 months but did not demonstrate this emotional cross-modal transfer ability in younger infants. These results suggested that the ability to amodally (i.e., independently of the sensorial modality) recognize happiness could emerge between 4 and 6 months of age.

Despite these early emotional skill developments, the ability to understand emotions improves considerably with age throughout childhood and (pre)adolescence (for reviews, see Gosselin, 2005; Harris, de Rosnay, & Pons, 2016; Pons et al., 2015; Pons, Harris, & de Rosnay, 2004). Indeed, as of age 2 years, children are able to make the first verbal categorization of certain primary emotions. Thereafter, this categorization continues to develop, and children become able to categorize an increasing number of emotions. Children will first categorize emotions into two types based on positive (pleasant) and negative (unpleasant) valences, and will only later reach a categorization of specific emotions as in adults (Widen & Russell, 2008, 2015). The development of children's emotion understanding seems to depend on the presentation used (e.g., stories, films, faces) and the response modalities (e.g., matching, label, etc.; Nelson & Russell, 2011; Vicari, Reilly, Pasqualetti, Vizzotto, & Caltagirone, 2000). It seems to also depend on the perceptive modality, with a recognition rate similar to that of adults for visual stimuli as of age 11, while for auditory stimuli this rate increases even more during adolescence (Chronaki, Hadwin, Garner, Maurage, & Sonuga-Barke, 2015).

However, a general pattern of development emerges from the various studies: Using verbal tasks, the identification of happiness is already well established at age 3; sadness, anger, and fear at around age 6; and the identification of surprise and disgust between the ages of 6 and 10 (Boyatzis, Chazan, & Ting, 1993; Camras & Allison, 1985; Durand, Gallay, Seigneuric, Robichon, & Baudouin, 2007; Gagnon, Gosselin, & Maassarani, 2014; Gosselin, 1995; Rodger, Vizioli, Ouyang, & Caldara, 2015; Widen & Russell, 2013). Moreover, emotional cross-modal studies in which participants are asked to judge the emotions with a varying degree of discordance between the affects expressed in a face and in a tone of voice demonstrate that adults (de Gelder & Vroomen, 2000; Gil, Hattouti, & Laval, 2016) and only 9 but not 5- or 7-year-old children (Gil et al., 2016) display a cross-modal effect with happy and sad expressions. These results suggest the existence of a bidirectional link between structures for detecting emotional expressions in both vision and audition increasing with age. In adult eye-tracking studies, intermodal stimulations revealed that both prosody and/or semantics orient more toward the congruent emotional face when faces and voices are presented simultaneously (Paulmann & Pell, 2010; Paulmann, Titone, & Pell, 2012; Rigoulot & Pell, 2012). Moreover, young adults (18–32 years) seem to recognize emotions better than older ones (60–87 years), suggesting a decline in emotional recognition with age (Sullivan, Ruffman, & Hutton, 2007). Nevertheless, this consensus is debatable due to the diversity of experimental designs and the requirement of language proficiency.

The emotional successive cross-modal transfer task presented to infants by Palama et al. (2018) allows us to study the development of emotional understanding ability without using language skills. Thus, the main goal of the present experiment is to examine how the spontaneous emotional successive transfer from voices to faces evolves in verbal participants. In this study, we applied the above paradigm for infants by Palama et al. (2018), to school-age children (ages 5, 8, and 10) and adults. Research with infants involves experiments that are not very long or complex, which can be conducted without giving instruction and without eliciting verbal responses. It is relatively unusual to use a visual preference paradigm to study visual recognition in verbal humans, however, this makes it possible to evaluate emotional recognition independently of the child's language abilities (access to the lexicon, labeling, etc.).

The secondary goal is therefore to explore the development of emotional face processing using eye-tracking and test whether the areas of interest (AOIs) on the face (eyes, mouth) are looked at differently depending on the age of the participant or emotions expressed. The study of the areas looked at during the production of a facial expression would make it possible to highlight more advanced discrimination skills. Indeed, when exploring an emotional face, some areas of the face may contain more useful information than others. Basic emotional facial expressions are produced with characteristic patterns of facial muscle movement. These patterns provide the perceptual basis for discriminating between different types of emotional expression (Ekman, Friesen, & Hager, 1978). We would like to explore if during cross-modal transfer all the areas of interest drive the same attention or if there are pattern specific areas as functions of the emotion expressed.

There is still little knowledge about the development of emotional face processing, despite the growing number of eye-tracking studies. One consistent result that can be found among the different studies is the higher proportion of fixation on the mouth and eye areas (e.g., Beaudry, Roy-Charland, Perron, Cormier, & Tapp, 2014; Eisenbarth & Alpers, 2011; Jack, Blais, Scheepers, Schyns, & Caldara, 2009; Schurgin et al., 2014; Vaidya, Jin, & Fellows, 2014). Among them, one eye-tracking study in infants shows that as of age 7 months, they already looked preferentially at the mouth for happy faces, whereas they oriented sooner toward the eye and eyebrow areas for angry and sad faces (Soussignan et al., 2018). Eye-tracking studies revealed that the eye area was looked at longer than the mouth for all basic emotions in children (de Wit, Falck-Ytter, & von Hofsten, 2008) and adults (Beaudry et al., 2014; Hernandez et al., 2009; Hunnius, de Wit, Vrins, & von Hofsten, 2011). It seems that a preferential looking behavior for the inner features of the face as a function of the emotion exists but it remains unclear and not always found (Blais, Fiset, Roy, Saumure Régimbald, & Gosselin, 2017; Jack et al., 2009; Vaidya et al., 2014). For example, Hunnius et al. (2011) suggested that the eye area is avoided in angry face perception, whereas others showed that the eyes are looked at longer in angry expressions (Eisenbarth & Alpers, 2011; Schurgin et al., 2014; Sullivan et al., 2007). Moreover, some studies suggest that happiness triggers more attention to the mouth area (Beaudry et al., 2014; Eisenbarth & Alpers, 2011; Hernandez et al., 2009; Schurgin et al., 2014).

Also, visual attention to faces seems correlated with better emotion recognition (Sullivan et al., 2007).

For a better understanding of cross-modal transfer in children and adults, we aimed at exploring if it would be present independently of the types of faces and voices used. For this purpose, we presented two stimulus conditions: One had the same stimuli as those used in Palama et al., 2018 (affective bursts and photographs of faces) as well as new ones (meaningless-speech sentences and 3D virtual faces). If a cross-modal transfer is found under different auditory or visual contexts, it would provide evidence for more a robust transfer ability.

In this study, we explored the development of the ability to represent emotion amodally. More precisely, we analyzed whether the perception of facial expressions is influenced by the voices previously heard. We investigated whether the face congruent to the voice was preferentially looked at, or whether as in 6-month-old infants, incongruency was preferred. More specifically, we examined whether potential visual preferences depended on attentional orientation toward specific areas of the face such as the eyes and/or mouth after auditory familiarization. We predicted that verbal children would be able to spontaneously match an emotional voice with the corresponding face but the mastering of this matching ability would increase with age. Moreover, we supposed that there would be a developmental pattern specific to each emotion: As the recognition of happiness comes earlier than that of anger, the transfer ability would be observed first with the happy voice and then with the angry voice.

Method

Participants

The sample comprised 80 participants consisting of 20 5-year-olds (7 females; 4.85 ± 0.58 years, range = 4 to 6 years), 20 8-year-olds (13 females; 8.15 ± 0.48 years, range = 7 to 9 years), 20 10-year-olds (8 females; 10.25 ± 0.55 years, range = 9 to 11 years), and 20 young adults (10 females; 21.80 ± 3.45 years, range = 18 to 32 years). In this experiment, each participant performed the two stimulus conditions: Condition 1: cross-modal transfer with affective bursts and photographs of faces, and Condition 2: cross-modal transfer with meaningless-speech voice and 3D virtual faces. We performed an a priori power analysis (calculated with G*Power 3.0.10) to determine the sample size required in order to have an effect size of .25, a power of .95, at $p < .05$, which would indicate that a total sample size of 76 subjects would be required. The Ethics Committees of the Faculty of Psychology and Educational Sciences at the University of Geneva gave approval for the study, and all parents and adult participants gave written informed consent for their children's or their own participation in the experiment.

Stimuli

Stimulus Condition 1: Affective bursts and photographs of faces. In stimulus Condition 1, auditory and visual stimuli were the same as those in the Palama et al. (2018) infant study. The emotional auditory stimuli of happiness, anger, and neutral came from the "Montreal Affective Voice" database (Belin, Fillion-Bilodeau, & Gosselin, 2008). They were expressive affective burst stimuli based on the emission of the vowel /a/ by one woman (ref: SF60). This auditory stimulus was a loop of a 1-s voice with a break of 1 s between each repetition for a total clip of 20 s.

The visual stimuli used were the happy and angry faces of a woman (ref: SF4) taken from the "The Karolinska Directed Emotional Faces (KDEF)" database (Lundqvist, Flykt, & Öhman, 1998; cf. Figure 1). These 14 × 10 cm, black-and-white photographs were presented in pairs on a medium gray background (RGB 100, 100, 100).

Stimulus Condition 2: Meaningless-speech voices and 3D virtual faces. The emotional auditory stimuli of happiness, anger, neutral came from the GEMEP database (Bänziger, Mortillaro, & Scherer, 2012). They were the repetition of two meaningless-speech sentences (linguistic phoneme sequences without semantic content; "nekal ibam soud molen" and "koun se mina lod belam") pronounced by three different actresses (ref: 2, 7, and 9) for a total clip of 20 s.

The emotional visual stimuli were 3D virtual happy and angry faces created with the Facial Action Coding System (FACS) FACSGen (Roesch et al., 2011; Figure 2). This software allows for the creation of highly standardized realistic synthetic 3D facial stimuli based on the FACSGen (Ekman et al., 1978). The face selected is a female face called "Tanja," to which we applied a gamma correction of 1.7. In order to make sure that the emotional expressions created were recognized as intended emotional faces (angry or happy), the stimuli were validated in a pilot study conducted with adult participants (see Appendix). These faces were 18 × 12 cm, in color, and were presented in pairs on a black background.

Experimental Procedure and Conditions

The children were tested directly in a quiet room at their school in Geneva. The adult participants were students recruited at the University of Geneva and tested in the Sensory-Motor, Affective and Social Lab. The participants were seated at a distance of 60 cm from the screen; at this distance, visual stimuli were 13.1° × 9.5° for Condition 1 and 16.7° × 11.3° for Condition 2 of visual angle. The stimulus display screen measured 47.5 cm × 30 cm with a spatial resolution of 1680 × 1050

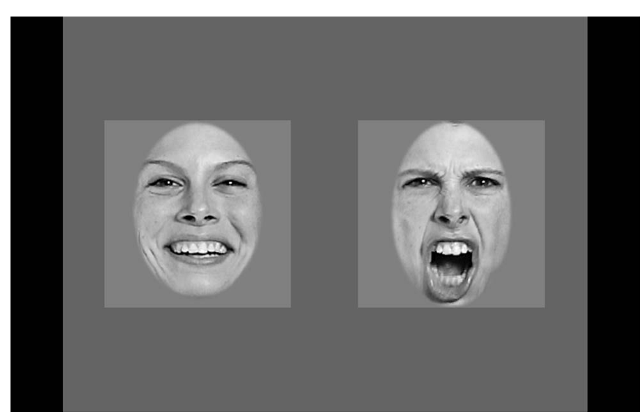

Figure 1. Visual stimulus Condition 1: happy face (left); angry face (right) with copyright permission from Lundqvist, Flykt, and Öhman (1998).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

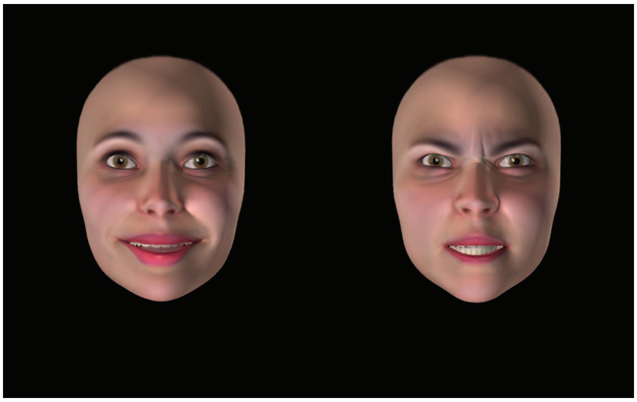

Figure 2. Visual stimulus Condition 2: happy face (left); angry face (right) with permission from Roesch et al. (2011). See the online article for the color version of this figure.

pixels. The gaze on visual stimuli was recorded with an eye-tracker SMI RED 250 (SensoMotoric Instruments GmbH, Teltow, Germany). The information provided before starting the experiment was that they would be listening to voices and watching faces. The only instruction was to look at the screen. The experiment started with a 5-point calibration phase with the eye-tracker, an animated image at five different locations covering the whole surface of the screen. This phase was repeated until a satisfactory calibration (less than 2° of deviation on the x and y axes) was achieved.

The experiment was composed of two stimulus conditions: Each participant performed the two stimulus conditions, and their order of presentation was randomized between participants. Between the two stimulus conditions we proceeded again to the calibration phase. The presentation of the 16 trials (sequences of audio-visual transfer) lasted 8 minutes for each participant. This experimental design was first used with infants, thus the stimulus presentation was longer but there were fewer stimuli presented than in typical adult paradigms. In this experiment, each trial consisted of exposure to a voice condition: no voice, neutral voice, and emotional happy or angry prosody for 20 s accompanied by a black display screen, for an auditory familiarization phase. Afterward, a happy and an angry face were presented side by side for 10 s during the visual test phase, in order to explore visual face processing carefully. The side the happy and angry faces were presented on was counterbalanced for each voice. Each condition consisted of 8 trials in this order. First, in order to obtain a baseline of the spontaneous preferences for emotional faces, the first 4 trials are 2 no-voice trials and 2 neutral-voice trials. The next 4 trials, the test trials, consisted of the presentation of the two emotional voices, angry and happy. Each trial was followed by the two emotional faces, laterally counterbalanced for each emotional voice, in succession (cf. Figure 3).

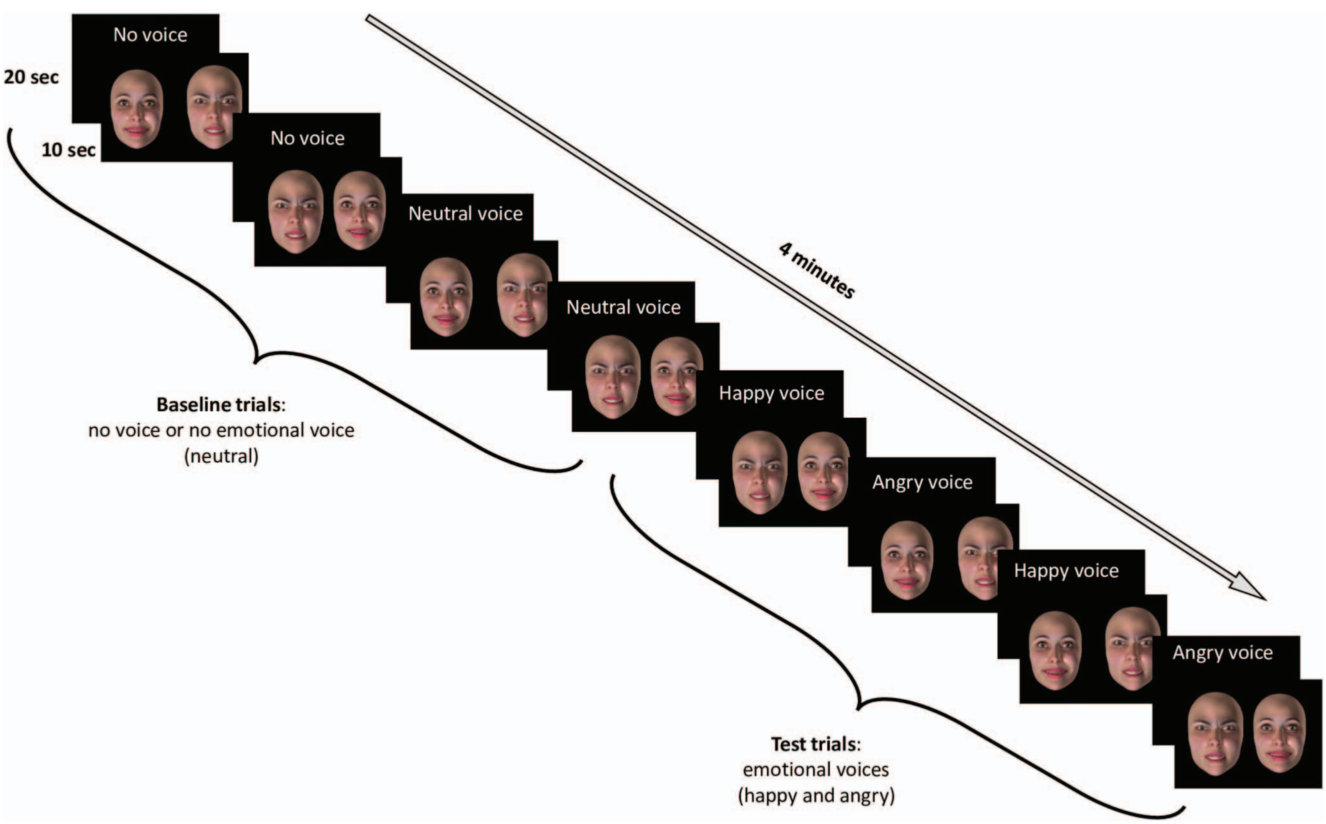

Figure 3. Schematic representation of the paradigm for stimulus Condition 2. The paradigm is exactly the same as that for stimulus Condition 1 but with different vocal and visual stimuli with permission from Roesch et al. (2011). See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Data Analysis

All the data were extracted by using Begaze SMI's analyzer software and can be found at https://osf.io/6r2gt/?view_only=7ca6993299f445bda1bbc598b64e1d24. In order to evaluate the development of the ocular behaviors according to the voice heard, we analyzed the looking time and first fixation for the AOIs: one general for the whole face and two specific ones for the eyes and the mouth (see Figure 4). In both conditions, these AOIs were of equal size for the expression of anger and happiness. The looking time was calculated by the net dwell time (length of time spent looking at the AOIs) in milliseconds (ms). The distribution of the first fixation on the happy or angry face was collected, and a fixation was determined as at least 80 ms in maximum 100 pixels of dispersion.

For each stimulus condition, we performed a repeated-measures analysis of variance (ANOVA) on the whole face and specific AOI looking times for the baseline condition (no voice and neutral voice) and the experimental condition (angry and happy voices). For the analysis, we split up the baseline from the experimental condition because it is only a control test to verify that there is no spontaneous preference for either face. In order to compare the two conditions' stimuli for the analysis of the emotional cross-modal transfer, the distribution of total looking time (DTLT) was calculated as the percentage of difference of looking time at happy faces (>0%) or angry faces (<0%); (looking time at happy faces − looking time at angry faces)/(looking time at happy faces + looking time at angry faces). A one-sample *t* test against chance was conducted with DTLT, to determine a looking preference for the emotional faces significantly greater than chance level (0%): a positive for happy faces and negative for angry ones. A repeated-measures ANOVA was performed on the rate of first fixation on each emotional face. Statistical analyses were conducted using Statistica 13. The significance threshold was .05. A Bonferroni test was performed to determine significant differences. For predicted interactions between emotional voices and emotional faces, a planned comparison was used to determine the preferential looking time at the congruent face or AOIs with the voice. Effect sizes are given in partial eta-squared (η_p^2) for ANOVAs.

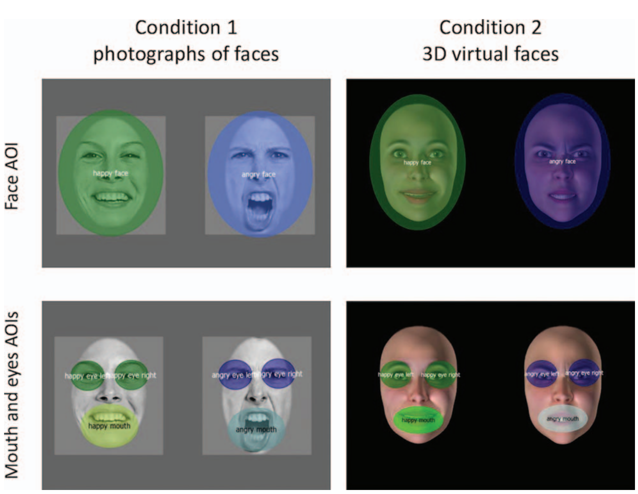

Figure 4. Areas of interest representing the whole face, eyes, and mouth. The happy face (left) and angry face (right): AOIs of the face, the eyes, and the mouth areas for Condition 1 with copyright permission from Lundqvist, Flykt, and Öhman (1998) and Condition 2 with permission from Roesch et al. (2011). See the online article for the color version of this figure.

Results

Baseline Analysis: Looking Time at Faces in No Voice and Neutral Voice Conditions

We analyzed the results of the baseline condition for the looking times at faces after the no voice or neutral voice condition as a function of age group for each stimulus condition (Conditions 1 and 2). A repeated-measures ANOVA was performed on looking times at faces with emotional faces (happy, angry) and voice condition (no voice, neutral voice) as within-subject factors and age (5, 8, 10 years and adults) as a between-subjects factor. The detailed results of the AOI analysis can be found in the Appendix.

Condition 1: Affective bursts and photographs of faces. A main effect of age, $F(3, 76) = 8.25$, $p < .001$, $\eta_p^2 = .246$, was revealed. Post hoc analyses using Bonferroni indicated that the 5-year-olds (3589 ± 131 ms) looked at faces for a shorter duration than the 10-year-olds (4259 ± 131; $p = .003$) and the adults (4465 ± 131; $p < .001$). A main effect of voice condition, $F(1, 76) = 19.83$, $p < .001$, $\eta_p^2 = .207$, was revealed: There were greater looking times after no voice (4269 ± 55 ms) than after the neutral voice (3895 ± 95 ms). The main effect of emotional face was not significant, $F(1, 76) = 1.46$, $p = .231$, $\eta_p^2 = .019$.

We observed a significant interaction between age and emotional faces, $F(3, 76) = 4.02$, $p = .010$, $\eta_p^2 = .137$. Post hoc analyses using Bonferroni indicated that the happy face in the 5-year-olds was looked at for a shorter time than the happy face in the adults, $p < .001$, or the angry face in the 10-year-olds, $p < .001$. However, in each age group, no difference in looking time between the emotional faces was found.

Condition 2: Meaningless-speech voices and 3D virtual faces. A main effect of age, $F(3, 76) = 10.22$, $p < .001$, $\eta_p^2 = .287$, was revealed. Post hoc analyses using Bonferroni indicated that the 5-year-olds (3838 ± 104 ms) had shorter looking times in general than the 8-year-olds (4268 ± 104; $p = .027$), 10-year-olds (4348 ± 104 ms; $p = .005$) and adults (4642 ± 104 ms; $p < .001$). A main effect of voice condition, $F(1, 76) = 16.99$, $p < .001$, $\eta_p^2 = .182$, was revealed: a greater looking time after no voice (4419 ± 47 ms) than after the neutral voice (4129 ± 75 ms). The main effect of emotional face was significant, $F(1, 76) = 6.95$, $p = .01$, $\eta_p^2 = .084$, suggesting that the happy face (4554 ± 122 ms) was looked at longer than the angry face (3994 ± 114 ms).

These main effects were qualified by an interaction between voice condition and emotional faces, $F(1, 76) = 4.822$, $p = .031$, $\eta_p^2 = .059$. Post hoc analyses using Bonferroni indicated that after the neutral voice, the happy face was looked at longer than the angry face ($p = .001$). Moreover, we observe a significant interaction between age and emotional faces, $F(3, 76) = 4.21$, $p = .008$, $\eta_p^2 = .143$. Post hoc analyses using Bonferroni indicated that the 8-year-olds looked longer at the happy face than the 5-year-olds; the adults looked longer at the happy face than the 5-year-olds looked at both emotional faces, and more than the 8-year-olds, 10-year-olds, and adults looked at the angry face, all *p*s < .05. We also observe a significant interaction between age and voice condition, $F(3, 76) = 4.97$, $p = .003$, $\eta_p^2 = .164$. Post hoc analyses using Bonferroni indicated that after the neutral voice, the 5-year-olds looked at the face for a shorter time than all other age and voice conditions, all *p*s <.05.

Analysis of the Emotional Cross-Modal Transfer

To respond to our hypothesis of the experimental conditions, we analyzed the DTLT and the direction of the first fixations to each emotional face after emotional voice presentation as a function of age group. Indeed, we were interested in examining if one of the emotional faces was looked at longer and/or faster than the other as a function of the voices presented. We have explored if there is greater DTLT and/or more first fixations for the congruent or the incongruent face in both stimulus conditions. For more information, the detailed results of the looking time for the face and AOI analysis can be found in the Appendix.

DTLT. A repeated-measures ANOVA was performed on DTLT at each emotional face with emotional voice (happy, angry) and stimulus condition (1, 2) as within-subject factors and age (5, 8, 10 years and adults) as a between-subjects factor. A repeated measures ANOVA was performed on DTLT at each emotional face with AOIs (mouth, eyes), emotional voice (happy, angry), and stimulus condition (1, 2) as within-subject factors and age (5, 8, 10 years and adults) as a between-subjects factor. A positive result represents a percentage of looking time favoring the happy face and a negative result represents the percentage of looking time favoring the angry face.

DTLT at emotional faces. The main effect of stimulus condition was not significant, $F(1, 76) = 1.16$, $p = .285$, $\eta_p^2 = .015$, and did not interact with any other factor, all *p*s > .19. The main effect of age was not significant, $F(3, 76) = 2.06$, $p = .119$, $\eta_p^2 = .074$. A main effect of emotional voice $F(1, 76) = 40.56$, $p < .001$, $\eta_p^2 = .378$, was revealed; the happy voice had a more positive DTLT than the angry voice. After a one-sample *t* test against chance level (0%) with the happy voice, the congruent happy face was looked at longer (18% ± 3%) than chance, $t(79) = 5.22$, $p < .001$, whereas with the angry voice it was the congruent angry face that was looked at longer (−19% ± 4%) than chance, $t(79) = -4.85$, $p < .001$ (cf. Figure 5).

This main effect was qualified by an interaction between age and emotional voice, $F(3, 76) = 5.19$, $p = .003$, $\eta_p^2 = .17$, (cf. Figure 6). After a one-sample t-test against chance level (0%), in 5-year-olds after the happy voice (4% ± 4%) and after the angry voice (−5% ± 5%), and in 8-year-olds after the angry voice (−6% ± 10%), none of the emotional faces was looked at longer than at chance level (0%), all $p > .29$. In the 8-year-olds, only after the happy voice, was the happy face looked at (20% ± 9%) more than at chance level (0%), $t(19) = 2.17$, $p = .043$. In the 10-year-olds, after the happy voice, the happy face was looked at longer (17% ± 5%) than at chance level (0%), $t(19) = 3.20$, $p = .005$, whereas after the angry voice it was the angry face that was looked at longer (−26% ± 6%) than chance, $t(19) = -4.08$, $p < .001$. In the adults, after the happy voice, the happy face was looked at longer (31% ± 7%) than at chance level (0%), $t(19) = 4.26$, $p < .001$, whereas after the angry voice it was the angry face that was looked at longer (−39% ± 7%), $t(19) = -5.46$, $p < .001$. All other interactions were non-significant.

A graph of the individual data is presented in Figure 7. Each data point represents the DTLT for one participant for both stimulus conditions, circle (Condition 1) or cross (Condition 2). The individual data points show that although the mean of the participants has higher DTLT for the congruent face, some have higher DTLT for the incongruent face. Moreover, DTLT at real faces (Condition 1) and virtual faces (Condition 2) were positively correlated, after the happy voice, $r(78) = .476$, $p < .05$, and after the angry voice $r(78) = .495$, $p < .05$.

DTLT at emotional AOIs. The main effect of stimulus condition was not significant, $F(1, 76) = 0.03$, $p = .87$, $\eta_p^2 = .00$. A main effect of age, $F(3, 76) = 2.89$, $p = .041$, $\eta_p^2 = .102$, was found. After a one-sample *t* test against chance level (0%), only the 10-year-olds looked at the angry AOIs (−11% ± 3%) more than chance. A main effect of emotional voice, $F(1, 76) = 31.99$, $p < .001$, $\eta_p^2 = .296$, was revealed: After the happy voice there were more positive DTLT at AOIs than after the angry voice. After a one-sample *t* test against chance level (0%), after the happy voice, the congruent happy AOIs were looked at longer (18% ± 4%) than chance, $t(79) = 4.62$, $p < .001$, whereas after the angry voice it was the congruent angry AOIs that were looked at longer (−20% ± 4%) than chance, $t(79) = -4.44$, $p < .001$.

These main effects were qualified by the interaction between age and emotional voice, $F(3, 76) = 31.99$, $p < .001$, $\eta_p^2 = .29$. After a one-sample *t* test against chance level (0%), in 5-year-olds with the happy (10% ± 6%) and angry voice (−2% ± 6%), and in 8-year-olds after the happy (17% ± 10%) and angry voice (−6% ± 10%), none of the emotional AOIs was looked at longer than at chance level (0%), all *p*s > .11. After the happy voice, the 10-year-olds tended to look at the happy AOIs (13% ± 6%) more than at chance level, $t(19) = 1.99$, $p = .060$, whereas after the angry voice it was the angry faces that were looked at longer (−35% ± 6%) than at chance level (0%), $t(19) = -5.37$, $p <$

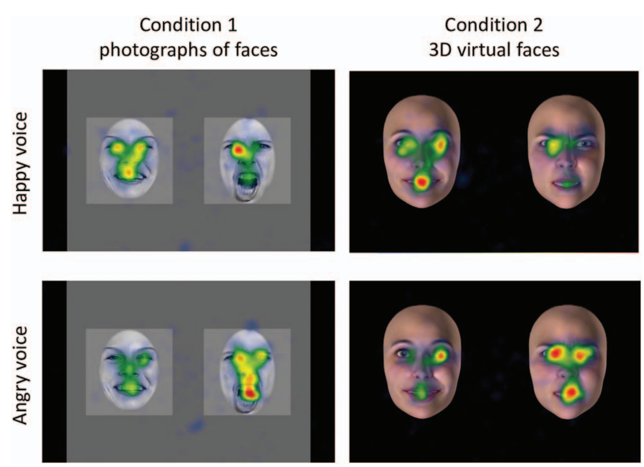

Figure 5. Heat map: all participants' gaze patterns over the stimulus image visualized as a colored map, blue (light grey) reprensent the minimum to red (dark grey) reprensent the maximum Fixation Time Average in ms. Visual stimulus Condition 1 with copyright persmission from Lundqvist, Flykt, and Öhman (1998); visual stimulus of Condition 2 with permission from Roesch et al. (2011). See the online article for the color version of this figure.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

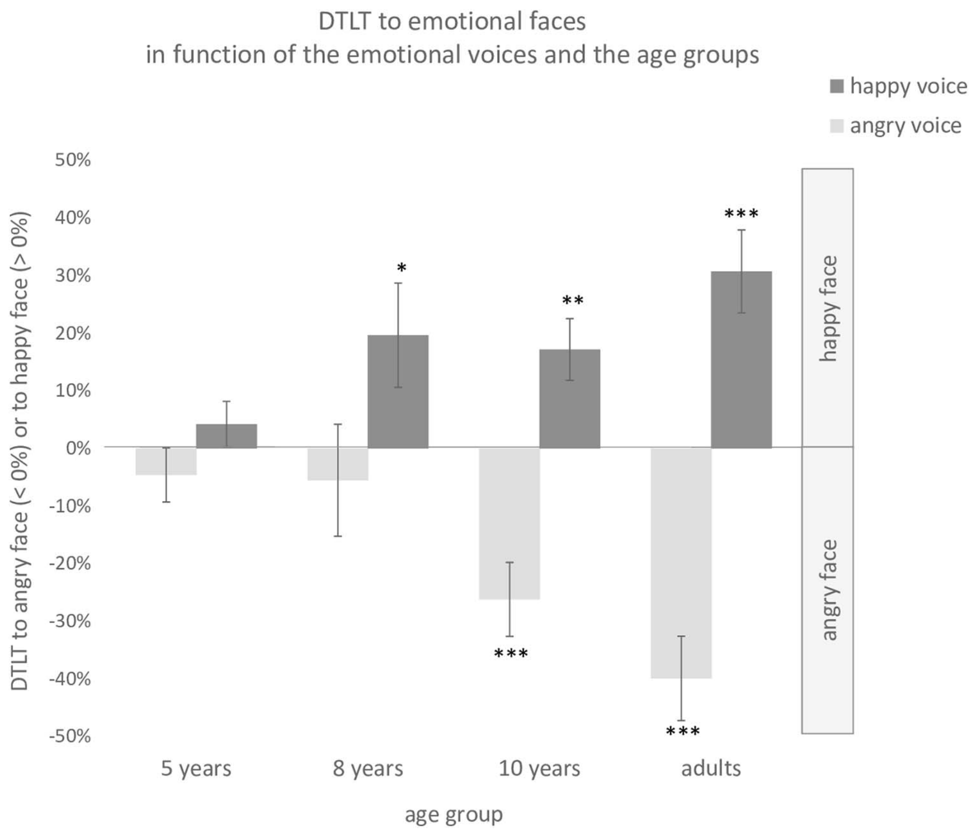

Figure 6. DTLT at happy (>0) or angry (<0) faces: as a function of voices, happy (dark gray) or angry (light gray), and ages (5, 8, 10 years and adults). * $p < .05$; ** $p < .01$; *** $p < .001$.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

.001. In the adults, after the happy voice, the happy faces are looked at longer (34% ± 8%) than at chance level (0%), $t(19) = 4.41$, $p < .001$, whereas after the angry voice it is the angry faces that are looked at longer (−38% ± 10%) than at chance level (0%), $t(19) = -3.85$, $p = .001$.

Another interaction was also significant: the interaction between AOIs and the stimulus condition, $F(1, 76) = 13.15$, $p < .001$, $\eta_p^2 = .148$. After a one-sample *t* test against chance level (0%), the real angry mouth is looked at longer (−7% ± 3%) than chance. None of the real eyes (6% ± 3%), or the virtual mouths (2% ± 3%) or eyes (−4% ± 3%), was looked at longer than chance, all *p*s >.088. All other interactions were nonsignificant.

First fixations to faces. A repeated-measures ANOVA was performed on the rate of first fixation [−2 to 2] on each emotional face with voice condition (no voice, neutral, happy, angry) and stimulus condition (1, 2) as within-subject factors and age (5, 8, 10 years and adults) as a between-subjects factor. A positive result [0–2] represented the rate of first fixation on the happy face, and a negative result [−2–0] represented the rate of first fixation on the angry face.

A main effect of stimulus condition, $F(1, 76) = 5.33$, $p = .02$, $\eta_p^2 = .066$, was revealed. After a one-sample *t* test against chance level (0%), the real angry face triggered the first fixation more (−0.18 ± 0.07) than chance, $t(79) = -2.46$, $p = .016$. All other main effects and interactions were nonsignificant.

Discussion

The main objective of this study was to assess children's and adults' eye behaviors in response to cross-modal emotional stimuli, and to explore whether this ability is dependent on age and the emotion expressed. The second objective was to determine whether the mouth and eye areas of the face are looked at differently as a function of the voices heard before. Through an emotional cross-modal transfer task (audio-visual), we used an eye-tracking analysis to evaluate the developmental evolution of looking behaviors for emotional faces in children age 5, 8, and 10, and young adults. Moreover, for a better understanding of cross-modal transfer performance in children and adults, we explored if this performance would be independent of the types of faces and voices used. To this purpose, cross-modal transfer in all participants was tested in two stimulus conditions. The first condition was cross-modal transfer with affective bursts and photographs of faces (the same stimuli used in Palama et al., 2018), and the second condition was cross-modal transfer with meaningless-speech sentences and 3D virtual faces.

In order to affirm that the emotional voices have an impact on the looking times of emotional faces, it is important to test the spontaneous preferences for faces and AOIs, that is after no voice or after a neutral one (baseline conditions). Results show similar patterns in both baseline conditions. With age, there is more

Figure 7. Individual DTLT at happy (>0) or angry (<0) faces: as a function of voices, happy (dark gray) or angry (light gray), and ages (5, 8, 10 years and adults). The bars in the background represents the mean DTLT to happy faces in dark gray and to angry faces in light grey.

attention on the faces and the internal features of the face (mouth and eyes AOIs). Moreover, as expected, after the no-voice condition, no significant difference between the time spent looking at the angry and happy faces was found for any age or stimulus condition. This result revealed no significant spontaneous visual preference for one of the emotional faces. After the neutral voice conditions, the results were different as a function of the stimulus condition presented. In stimulus Condition 1, there was no significant difference of looking time between the happy and the angry face photographs. However, in stimulus Condition 2, the 3D virtual happy faces were looked at longer than the angry ones, particularly in the adults. They seemed to search more for a matching face to the voice than the children did. In stimulus Condition 2, the voice was meaningless-speech sentences portrayed by three women, while in stimulus Condition 1, the voice was affective bursts from a woman pronouncing the vowel sound /a/. Thus, this result suggests that female voices that express meaningless speech, even in a neutral tone, are more associated with the happy face than the angry face. However, a female voice that pronounces only the vowel /a/ with a neutral tone of voice was not associated with any of the emotional faces, suggesting a neutral triggering. This result indicated that a voice that pronounces pseudowords even with a neutral tone is considered positive, while a voice that pronounces just one phoneme with a neutral tone is considered neutral. Producing speech composed of pseudowords seems to be more associated with positive emotion than neutral emotion.

Furthermore, concerning the AOIs looked at after the baseline condition, another effect found was that the angry mouth was looked at longer than the happy one in Condition 1. In stimulus Condition 2, as of age 8, the eyes were looked at longer than the mouth, thus confirming previous findings (Beaudry et al., 2014; Hunnius et al., 2011).

If we compare the two different baseline conditions used to determine the spontaneous visual preference, the no-voice condition didn't show any evidence that angry and happy faces were looked at differently. The neutral voice seems to reveal more ambiguous results. Even though the voice is proven to be neutral, it could modify spontaneous visual preferences.

Concerning our main objective, which is to explore the development of emotional cross-modal transfer, emotional voices (happy and angry) seemed to drive attention to the face and AOIs (mouth and eyes) that were more congruent to the voice in both stimulus conditions (1 and 2). As predicted, this result suggested that happy and angry emotions were spontaneously recognized in an amodal way independently of these conditions, suggesting a robust transfer ability. Preference for the congruent expression in the children and young adults was in accord with some previous studies using a simultaneous intermodal task in infants (Godard et al., 2016; Kahana-Kalman & Walker-Andrews, 2001; Montague & Walker-Andrews, 2002; Soken & Pick, 1992; Vaillant-Molina et al., 2013; Walker, 1982; Zieber et al., 2014) and adults (Paulmann & Pell, 2010; Paulmann et al., 2012; Rigoulot & Pell, 2012). However, these results were in contrast to those observed in 3-month-old (Montague & Walker-Andrews, 2002) and 6-month-old infants in a previous experiment using the same paradigm (Palama et al., 2018, 2020). This leads us to believe that there may be a change in development from a preference for the "incongruent" novel expression to a preference for the "congruent" familiar expression. Indeed, in accordance with the theory of the violation of expectation (Wang, Baillargeon, & Brueckner, 2004), some studies showed that infants and toddlers prefer to focus their attention on the incongruent emotional scene (Hepach & Westermann, 2013; Reschke, Walle, Flom, & Guenther, 2017; Skerry & Spelke, 2014).

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

However, as expected, the mastering of this matching ability increases with age and there is a developmental pattern specific to each emotion. Indeed, the preference for the congruent expression was significantly revealed only as of age 8 with the happy voice and as of age 10 with the angry voice. A discussion of these results using an infants' experimental design may be based on two interpretations.

First, in this experiment, children had no instruction except to look at the screen. Thus, younger children could understand the emotions expressed but didn't have a spontaneous preference observable for the congruent or the incongruent one during this stimulus presentation, so they explored both faces. With this nonverbal cross-modal transfer paradigm (Palama et al., 2018, 2020), the more familiar expression of happiness is transferred with an incongruency preference in 6-month-old infants while in this study, no significant preference has been highlighted in 5-year-olds, whereas as of age 8 children showed reversed preference in favor of congruence. Thus, 5 would be a pivotal age from the reaction to novelty to a more mature preference for congruence. This interpretation seems possible because studies have shown that children are already able to correctly label more than 90% of the facial expressions of happiness as of age 3 and anger as of age 4 (Widen, 2013). However, the recognition of emotional prosody using a free-labeling task does not exceed 50% in 5-to-7-year-old children (Nelson & Russell, 2011; Sauter, Panattoni, & Happé, 2013). Indeed, it is well known that these recognition skills are affected by the kind of task (e.g., matching, labialization, stories, faces, etc.), all of which involve verbal instructions (Nelson & Russell, 2011; Vicari et al., 2000). Moreover, Berman, Chambers, and Graham (2016) have demonstrated that 5-year-old children match the happy and the sad prosody with the corresponding emotional face if they are instructed to do so. In view of these considerations, we can assume that a cross-modal transfer could be found earlier in childhood, if instructions were given to children.

Second, even though the emotional facial expressions are correctly labeled for happy and angry before age 5 (Widen, 2013), the recognition ability is not totally acquired at that age. Indeed, recognition of emotion increases with age and studies have found that 11-year-olds' accuracy in facial expression recognition is similar to that of adults, whereas adult-level vocal expression recognition is achieved in later adolescence (Chronaki et al., 2015). In our study, the youngest children (5-year-olds) showed no significant preference for one of the facial expressions after both emotional voices. This result could suggest that they do not master emotion recognition and so are unable to match the facial expressions with the corresponding voice before age 8. Moreover, 8-year-old children are able to match the facial expression with the happy voice, but our study did not demonstrate this with the angry one. This result could be explained by the fact that the happy vocal expression is actually understood at this age and it is only with this voice that children are able to match the facial expression. This result is in line with studies that show that happiness is recognized first, and other emotions like anger are recognized later in development (Boyatzis et al., 1993; Camras & Allison, 1985; Durand et al., 2007; Gagnon et al., 2014; Gosselin, 1995; Rodger et al., 2015; Widen & Russell, 2013). The congruent transfer for happiness over anger in younger subjects is then consistent with the earlier development of understanding this emotion. However, it should be noted that some studies do not demonstrate a difference in the age of recognition between the angry and the happy voice, for example the forced choice task of Chronaki et al. (2015), and thus this difference of developmental recognition could be task specific. Moreover, this result is consistent with the study conducted by Gil et al. (2016), which shows that emotional cross-modal transfer is only observed as of 9 years of age. This suggests that a bidirectional link between structures for detecting emotional expressions in both vision and audition could emerge later in childhood.

The relatively late mastering of emotional cross-modal transfer could also have several causes. One of these causes could be that the two stimuli are presented successively and not simultaneously. Indeed, it is easier for children to associate two perceived elements simultaneously and more in line with reality. Indeed, children age 7 to 12 showed labeling accuracy of 100% for happiness, anger, or sadness if faces and voices are congruent (Shackman & Pollak, 2005). However, this paradigm with simultaneous presentation of stimuli does not allow us to know whether children base their answers on what they have seen or heard. Another cause could be that the visual stimuli used were static, while the emotional expressions are inherently dynamic (Caron, Caron, & Myers, 1985; Wilcox, Stubbs, Wheeler, & Alexander, 2013). Bänziger, Grandjean, and Scherer (2009) demonstrated that emotions presented in multimodal (audio-visual) ways are better recognized than emotions presented in unimodal ways. In addition, dynamic expressions appear to be more easily recognized than static facial expressions.

AOI analysis also revealed a preference in looking times at the congruent AOIs with voices. However, there is a delay in the age of appearance of the effect compared to the whole face. Indeed, it is only as of age 10 that we found evidence of the fact that children look more at the AOIs of the faces congruent with the angry voice and as of adulthood with the happy voice. This result is in accord with the fact that internal features are considered to become more critical for adulthood face expertise (Ellis, Shepherd, & Davies, 1979; Ge et al., 2008; Pascalis et al., 2011; Tanaka & Farah, 1993) and are looked at longer to process emotional faces.

Moreover, with AOI analysis, we found different results as a function of the stimulus condition. In stimulus Condition 1, the time to look at the mouth and eyes was not found to be different, in accordance with the study by Amso, Fitzgerald, Davidow, Gilhooly, and Tottenham (2010). However, more participants looked at the angry mouth than at the happy one. This result suggests that the angry mouth seems more attractive after emotional voices, similarly to what was found in 6-month-old infants (Palama et al., 2018) using the same stimuli. In stimulus Condition 2, more attention was paid to the eye area compared to the mouth in all age groups. This result is consistent with previous studies in infants (Hunnius et al., 2011; Peltola, Leppänen, Vogel-Farley, Hietanen, & Nelson, 2009; Soussignan et al., 2018), children (de Wit et al., 2008), and adults (Beaudry et al., 2014; Hernandez et al., 2009; Hunnius et al., 2011). These results also suggest that the mouth of the face in Condition 1 biases regular visual attention.

Regarding first fixation, there is only one effect, the main effect of stimulus condition, suggesting that in stimulus Condition 1, participants directed their first spontaneous attention toward the angry face regardless of the voice, whereas in stimulus Condition 2, no difference of first fixation was significant. This result could be explained by the possible threat expressed by these real face photographs, driving the first attention to it, while perhaps 3D virtual faces are identified as less threatening even if they express anger. Even though the real and therefore potentially more threatening face attracts the first fixations, emotional cross-modal transfer is observed with both stimulus conditions. Thus, this pattern shows a minor effect of the stimulus conditions that does not affect the emotional intermodal transfer abilities.

Thus, to clarify the interpretations of the results, additional studies are required. It would be interesting to evaluate the effects on a larger number of participants, but also to administer more trials and present other emotions such as sadness, fear, or disgust. Moreover, future studies should use an active task, in which children would be given specific instructions or be tested for their unimodal (visual and auditory) verbal identification of the stimuli presented.

To conclude, this experiment revealed that children as of age 8 and adults were spontaneously able to match the vocal expression and they oriented their attention toward the congruent facial expression in different stimulus conditions. This confirms that a cross-modal transfer of emotional information was used by participants in this paradigm, demonstrating understanding of the amodal nature of emotion. Future research needs to establish a link between emotion understanding ability evaluated with classical tasks and the cross-modal emotional transfer ability. From a clinical perspective, further research should validate if this is an appropriate task to evaluate emotion understanding in atypical populations that are unable to understand verbal instructions or produce verbal responses.

References

Amso, D., Fitzgerald, M., Davidow, J., Gilhooly, T., & Tottenham, N. (2010). Visual exploration strategies and the development of infants' facial emotion discrimination. *Frontiers in Psychology, 1,* 180. http://dx.doi.org/10.3389/fpsyg.2010.00180

Bänziger, T., Grandjean, D., & Scherer, K. R. (2009). Emotion recognition from expressions in face, voice, and body: The Multimodal Emotion Recognition Test (MERT). *Emotion, 9,* 691–704. http://dx.doi.org/10.1037/a0017088

Bänziger, T., Mortillaro, M., & Scherer, K. R. (2012). Introducing the Geneva Multimodal expression corpus for experimental research on emotion perception. *Emotion, 12,* 1161–1179. http://dx.doi.org/10.1037/a0025827

Bayet, L., Pascalis, O., & Gentaz, É. (2014). Le développement de la discrimination des expressions faciales émotionnelles chez les nourrissons dans la première année [The development of emotional facial expression discrimination by infants in the first year of life]. *L'Année Psychologique, 114,* 469–500. http://dx.doi.org/10.4074/S0003503314003030

Beaudry, O., Roy-Charland, A., Perron, M., Cormier, I., & Tapp, R. (2014). Featural processing in recognition of emotional facial expressions. *Cognition and Emotion, 28,* 416–432. http://dx.doi.org/10.1080/02699931.2013.833500

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. *Behavior Research Methods, 40,* 531–539. http://dx.doi.org/10.3758/BRM.40.2.531

Berman, J. M. J., Chambers, C. G., & Graham, S. A. (2016). Preschoolers' real-time coordination of vocal and facial emotional information. *Journal of Experimental Child Psychology, 142,* 391–399. http://dx.doi.org/10.1016/j.jecp.2015.09.014

Birch, H. G., & Lefford, A. (1963). Intersensory development in children. *Monographs of the Society for Research in Child Development, 28,* 1–48. http://dx.doi.org/10.2307/1165681

Blais, C., Fiset, D., Roy, C., Saumure Régimbald, C., & Gosselin, F. (2017). Eye fixation patterns for categorizing static and dynamic facial expressions. *Emotion, 17,* 1107–1119. http://dx.doi.org/10.1037/emo0000283

Boersma, P., & Van Heuven, V. (2018). *Praat: Doing phonetics by computer* (Version 6.0.20) [Computer program]. Retrieved from http://www.praat.org/

Boyatzis, C. J., Chazan, E., & Ting, C. Z. (1993). Preschool children's decoding of facial emotions. *The Journal of Genetic Psychology, 154,* 375–382. http://dx.doi.org/10.1080/00221325.1993.10532190

Bryant, P. (1974). *Perception and understanding in young children: An experimental approach*. New York, NY: Routledge.

Camras, L. A., & Allison, K. (1985). Children's understanding of emotional facial expressions and verbal labels. *Journal of Nonverbal Behavior, 9,* 84–94. http://dx.doi.org/10.1007/BF00987140

Caron, R. F., Caron, A. J., & Myers, R. S. (1985). Do infants see emotional expressions in static faces? *Child Development, 56,* 1552–1560. http://dx.doi.org/10.2307/1130474

Castro, V. L., Cheng, Y., Halberstadt, A. G., & Grühn, D. (2016). EUReKA! A conceptual model of emotion understanding. *Emotion Review, 8,* 258–268. http://dx.doi.org/10.1177/1754073915580601

Chronaki, G., Hadwin, J. A., Garner, M., Maurage, P., & Sonuga-Barke, E. J. S. (2015). The development of emotion recognition from facial expressions and non-linguistic vocalizations during childhood. *British Journal of Developmental Psychology, 33,* 218–236. http://dx.doi.org/10.1111/bjdp.12075

de Gelder, B., & Vroomen, J. (2000). The perception of emotions by ear and by eye. *Cognition and Emotion, 14,* 289–311. http://dx.doi.org/10.1080/026999300378824

de Wit, T. C. J., Falck-Ytter, T., & von Hofsten, C. (2008). Young children with Autism Spectrum Disorder look differently at positive versus negative emotional faces. *Research in Autism Spectrum Disorders, 2,* 651–659. http://dx.doi.org/10.1016/j.rasd.2008.01.004

Durand, K., Gallay, M., Seigneuric, A., Robichon, F., & Baudouin, J.-Y. (2007). The development of facial emotion recognition: The role of configural information. *Journal of Experimental Child Psychology, 97,* 14–27. http://dx.doi.org/10.1016/j.jecp.2006.12.001

Eisenbarth, H., & Alpers, G. W. (2011). Happy mouth and sad eyes: Scanning emotional facial expressions. *Emotion, 11,* 860–865. http://dx.doi.org/10.1037/a0022758

Ekman, P., Friesen, W. V., & Hager, J. C. (1978). *Facial action coding system (FACS): A technique for the measurement of facial action*. Palo Alto, CA: Consulting Psychologists Press.

Ellis, H. D., Shepherd, J. W., & Davies, G. M. (1979). Identification of familiar and unfamiliar faces from internal and external features: Some implications for theories of face recognition. *Perception, 8,* 431–439. http://dx.doi.org/10.1068/p080431

Gagnon, M., Gosselin, P., & Maassarani, R. (2014). Children's ability to recognize emotions from partial and complete facial expressions. *The Journal of Genetic Psychology, 175,* 416–430. http://dx.doi.org/10.1080/00221325.2014.941322

Ge, L., Anzures, G., Wang, Z., Kelly, D. J., Pascalis, O., Quinn, P. C., . . . Lee, K. (2008). An inner face advantage in children's recognition of familiar peers. *Journal of Experimental Child Psychology, 101,* 124–136. http://dx.doi.org/10.1016/j.jecp.2008.05.006

Gil, S., Hattouti, J., & Laval, V. (2016). How children use emotional prosody: Crossmodal emotional integration? *Developmental Psychology, 52,* 1064–1072. http://dx.doi.org/10.1037/dev0000121

Godard, O., Baudouin, J.-Y., Schaal, B., & Durand, K. (2016). Affective matching of odors and facial expressions in infants: Shifting patterns between 3 and 7 months. *Developmental Science, 19,* 155–163. http://dx.doi.org/10.1111/desc.12292

Gosselin, P. (1995). Le développement de la reconnaissance des expressions faciales des émotions chez l'enfant [The development of the

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

recognition of emotional facial expressions in children]. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement, 27,* 107–119. http://dx.doi.org/10.1037/008-400X.27.1.107

Gosselin, P. (2005). Le décodage de l'expression faciale des émotions au cours de l'enfance [The emotional decoding of facial expressions during the duration of childhood]. *Canadian Psychology/Psychologie canadienne, 46,* 126–138. http://dx.doi.org/10.1037/h0087016

Harris, P. L., de Rosnay, M., & Pons, F. (2016). Understanding emotion. In L. Feldman Barrett, M. Lewis, & J. M. Haviland-Jones (Eds.), *Handbook of emotions* (pp. 293–306). New York, NY: Guilford Press.

Hepach, R., & Westermann, G. (2013). Infants' sensitivity to the congruence of others' emotions and actions. *Journal of Experimental Child Psychology, 115,* 16–29. http://dx.doi.org/10.1016/j.jecp.2012.12.013

Hernandez, N., Metzger, A., Magné, R., Bonnet-Brilhault, F., Roux, S., Barthelemy, C., & Martineau, J. (2009). Exploration of core features of a human face by healthy and autistic adults analyzed by visual scanning. *Neuropsychologia, 47,* 1004–1012. http://dx.doi.org/10.1016/j.neuropsychologia.2008.10.023

Hunnius, S., de Wit, T. C. J., Vrins, S., & von Hofsten, C. (2011). Facing threat: Infants' and adults' visual scanning of faces with neutral, happy, sad, angry, and fearful emotional expressions. *Cognition and Emotion, 25,* 193–205. http://dx.doi.org/10.1080/15298861003771189

Jack, R. E., Blais, C., Scheepers, C., Schyns, P. G., & Caldara, R. (2009). Cultural confusions show that facial expressions are not universal. *Current Biology, 19,* 1543–1548. http://dx.doi.org/10.1016/j.cub.2009.07.051

Kahana-Kalman, R., & Walker-Andrews, A. S. (2001). The role of person familiarity in young infants' perception of emotional expressions. *Child Development, 72,* 352–369. http://dx.doi.org/10.1111/1467-8624.00283

Lundqvist, D., Flykt, A., & Öhman, A. (1998). The Karolinska Directed Emotional Faces—KDEF [CD-ROM]. Stockholm, Sweden: Karolinska Institute, Department of Clinical Neuroscience, Psychology Section.

Montague, D. R., & Walker-Andrews, A. S. (2002). Mothers, fathers, and infants: The role of person familiarity and parental involvement in infants' perception of emotion expressions. *Child Development, 73,* 1339–1352. http://dx.doi.org/10.1111/1467-8624.00475

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

Palama, A., Malsert, J., & Gentaz, E. (2018). Are 6-month-old human infants able to transfer emotional information (happy or angry) from voices to faces? An eye-tracking study. *PLoS ONE, 13,* e0194579. http://dx.doi.org/10.1371/journal.pone.0194579

Palama, A., Malsert, J., & Gentaz, É. (2020). *The cross-modal transfer of emotional information (happy or angry) from voices to faces in 2, 4 and 6 month-old infants*. Manuscript submitted for publication.

Pascalis, O., de Martin de Viviés, X., Anzures, G., Quinn, P. C., Slater, A. M., Tanaka, J. W., & Lee, K. (2011). Development of face processing. *WIREs Cognitive Science, 2,* 666–675. http://dx.doi.org/10.1002/wcs.146

Paulmann, S., & Pell, M. D. (2010). Contextual influences of emotional speech prosody on face processing: How much is enough? *Cognitive, Affective & Behavioral Neuroscience, 10,* 230–242. http://dx.doi.org/10.3758/CABN.10.2.230

Paulmann, S., Titone, D., & Pell, M. D. (2012). How emotional prosody guides your way: Evidence from eye movements. *Speech Communication, 54,* 92–107. http://dx.doi.org/10.1016/j.specom.2011.07.004

Peltola, M. J., Leppänen, J. M., Vogel-Farley, V. K., Hietanen, J. K., & Nelson, C. A. (2009). Fearful faces but not fearful eyes alone delay attention disengagement in 7-month-old infants. *Emotion, 9,* 560–565. http://dx.doi.org/10.1037/a0015806

Pons, F., Guiménez-Dasi, M., Nives Sala, M., Molina, P., Tornare, E., & Anderson, B. (2015). Compréhension et régulation des émotions à l'école [Comprehension and regulation of emotions at school]. In M. Crahay & M. Dutrevis (Eds.), *Psychologie des apprentissages scolaires* (pp. 115–142). Ottignies-Louvain-la-Neuve, Belgium: De Boeck Supérieur.

Pons, F., Harris, P. L., & de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: Developmental periods and hierarchical organization. *European Journal of Developmental Psychology, 1,* 127–152. http://dx.doi.org/10.1080/17405620344000022

Reschke, P. J., Walle, E. A., Flom, R., & Guenther, D. (2017). Twelve-month-old infants' sensitivity to others' emotions following positive and negative events. *Infancy, 22,* 874–881. http://dx.doi.org/10.1111/infa.12193

Rigoulot, S., & Pell, M. D. (2012). Seeing emotion with your ears: Emotional prosody implicitly guides visual attention to faces. *PLoS ONE, 7,* e30740. http://dx.doi.org/10.1371/journal.pone.0030740

Rodger, H., Vizioli, L., Ouyang, X., & Caldara, R. (2015). Mapping the development of facial expression recognition. *Developmental Science, 18,* 926–939. http://dx.doi.org/10.1111/desc.12281

Roesch, E. B., Tamarit, L., Reveret, L., Grandjean, D., Sander, D., & Scherer, K. R. (2011). FACSGen: A tool to synthesize emotional facial expressions through systematic manipulation of facial action units. *Journal of Nonverbal Behavior, 35,* 1–16. http://dx.doi.org/10.1007/s10919-010-0095-9

Russell, J. A., & Dols, J. M. F. (2017). *The science of facial expression*. New York, NY: Oxford University Press. http://dx.doi.org/10.1093/acprof:oso/9780190613501.003.0024

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

Schurgin, M. W., Nelson, J., Iida, S., Ohira, H., Chiao, J. Y., & Franconeri, S. L. (2014). Eye movements during emotion recognition in faces. *Journal of Vision, 14*(13), 14. http://dx.doi.org/10.1167/14.13.14

Shackman, J. E., & Pollak, S. D. (2005). Experiential influences on multimodal perception of emotion. *Child Development, 76,* 1116–1126. http://dx.doi.org/10.1111/j.1467-8624.2005.00901.x

Skerry, A. E., & Spelke, E. S. (2014). Preverbal infants identify emotional reactions that are incongruent with goal outcomes. *Cognition, 130,* 204–216. http://dx.doi.org/10.1016/j.cognition.2013.11.002

Soken, N. H., & Pick, A. D. (1992). Intermodal perception of happy and angry expressive behaviors by seven-month-old infants. *Child Development, 63,* 787–795. http://dx.doi.org/10.2307/1131233

Soussignan, R., Dollion, N., Schaal, B., Durand, K., Reissland, N., & Baudouin, J.-Y. (2018). Mimicking emotions: How 3–12-month-old infants use the facial expressions and eyes of a model. *Cognition and Emotion, 32,* 827–842. http://dx.doi.org/10.1080/02699931.2017.1359015

Sullivan, S., Ruffman, T., & Hutton, S. B. (2007). Age differences in emotion recognition skills and the visual scanning of emotion faces. *The Journals of Gerontology: Series B, 62,* P53–P60. http://dx.doi.org/10.1093/geronb/62.1.P53

Tanaka, J. W., & Farah, M. J. (1993). Parts and wholes in face recognition. *The Quarterly Journal of Experimental Psychology Section A, 46,* 225–245. http://dx.doi.org/10.1080/14640749308401045

Vaidya, A. R., Jin, C., & Fellows, L. K. (2014). Eye spy: The predictive value of fixation patterns in detecting subtle and extreme emotions from faces. *Cognition, 133,* 443–456. http://dx.doi.org/10.1016/j.cognition.2014.07.004

Vaillant-Molina, M., Bahrick, L. E., & Flom, R. (2013). Young infants match facial and vocal emotional expressions of other infants. *Infancy, 18,* E97–E111. http://dx.doi.org/10.1111/infa.12017

Vicari, S., Reilly, J. S., Pasqualetti, P., Vizzotto, A., & Caltagirone, C. (2000). Recognition of facial expressions of emotions in school-age children: The intersection of perceptual and semantic categories. *Acta*

Paediatrica, 89, 836–845. http://dx.doi.org/10.1111/j.1651-2227.2000.tb00392.x

Walker, A. S. (1982). Intermodal perception of expressive behaviors by human infants. *Journal of Experimental Child Psychology, 33,* 514–535. http://dx.doi.org/10.1016/0022-0965(82)90063-7

Wang, S. H., Baillargeon, R., & Brueckner, L. (2004). Young infants' reasoning about hidden objects: Evidence from violation-of-expectation tasks with test trials only. *Cognition, 93,* 167–198. http://dx.doi.org/10.1016/j.cognition.2003.09.012

Widen, S. C. (2013). Children's interpretation of facial expressions: The long path from valence-based to specific discrete categories. *Emotion Review, 5,* 72–77. http://dx.doi.org/10.1177/1754073912451492

Widen, S. C., & Russell, J. A. (2008). Children acquire emotion categories gradually. *Cognitive Development, 23,* 291–312. http://dx.doi.org/10.1016/j.cogdev.2008.01.002

Widen, S. C., & Russell, J. A. (2013). Children's recognition of disgust in others. *Psychological Bulletin, 139,* 271–299. http://dx.doi.org/10.1037/a0031640

Widen, S. C., & Russell, J. A. (2015). Do dynamic facial expressions convey emotions to children better than do static ones? *Journal of Cognition and Development, 16,* 802–811. http://dx.doi.org/10.1080/15248372.2014.916295

Wilcox, T., Stubbs, J. A., Wheeler, L., & Alexander, G. M. (2013). Infants' scanning of dynamic faces during the first year. *Infant Behavior and Development, 36,* 513–516. http://dx.doi.org/10.1016/j.infbeh.2013.05.001

Zieber, N., Kangas, A., Hock, A., & Bhatt, R. S. (2014). The development of intermodal emotion perception from bodies and voices. *Journal of Experimental Child Psychology, 126,* 68–79. http://dx.doi.org/10.1016/j.jecp.2014.03.005

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

Appendix

Additional information and results

Visual Stimulus Condition 2 Creation and Validation Study

For each emotional face, we activated the action units (AUs) in accordance with Ekman et al., 1978; and Roesch et al., 2011. For the happy face, AUs 6, 12, 25, and 26 were activated, with an intensity of activation (in order of each AU) of 60, 60, 40, and 50 (i.e. maximum intensity is 100). For the angry face AU 4, 5, 7, 9, 16, 23, and 25 were activated, with an intensity of activation of 60 for each AU (cf. Table A1, Figure 2).

In order to make sure that the emotional faces created with FACSGen were recognized as intended emotional faces (angry or happy), the stimuli were validated in a pilot study with adult participants, 24 University of Geneva students (14 females; mean age= 21.04 ± 2.8 years). Each participant was instructed to classify the faces in a 7-alternative, forced-choice task. The choices were the following: happiness, anger, sadness, disgust, surprise, fear, and a black field to allow participants to suggest another emotion. The stimuli and the list of choices were presented in a randomized order. Stimuli were recognized with an accuracy of 92% (N = 22/24) for both emotional expressions, much higher than the chance level of 14.29% (100/7).

Table A1
Action Units (AUs) and Intensity Used to Create Happy and Angry Faces With FACSGen

AU number	AU FACS name	Intensity of activation maximum = 100	
		Happy	Angry
4	Brow Lowerer	0	60
5	Upper Lid Raiser	0	60
6	Cheek Raiser	60	0
7	Lid Tightener	0	60
9	Nose Wrinkler	0	60
12	Lip Corner Puller	60	0
16	Lower Lip Depressor	0	60
23	Lip Tightener	0	60
25	Lip Part	40	60
26	Jaw Drop	50	0

Acoustic Analyses of the Vocal Stimuli

Full acoustic analyses of all the vocal stimuli performed with Praat software (Boersma & Van Heuven, 2018) are presented in Table A2. The volume of the auditory stimuli presented did not exceed 60 dBA.

General Looking Time Description

With all voice conditions (no voice, neutral, happy, or angry), in stimulus condition (1, 2) and in all age groups (5, 8, 10 years and adults), the mean of looking time at faces is about 8037/10000 ms, ranging from 7516 to 8838 ms. In all faces, gazes are directed toward the inner features of the faces, the eyes, nose, and mouth (cf. Figure A1).

Looking Time Analysis of Condition 1: Affective Bursts and Photographs of Faces

Baseline condition: Looking time at AOIs after no voice and neutral voice. We analyzed the results of the baseline condition for the looking times at AOIs (eyes, mouth) presented after the no-voice or neutral-voice condition as a function of age group. A repeated-measures ANOVA was performed on looking times at AOIs with emotional faces (happy, angry), AOIs (eyes, mouth), and voice condition (no voice, neutral voice) as within-subject factors and age (5, 8, 10 years and adults) as a between-subjects factor.

(*Appendix continues*)

Table A2
Acoustic Analyses of All the Vocal Stimuli Performed With Praat Software

Sound	Pitch				Intensity				Duration
	Average	Standard deviation	Minimum	Maximum	Average	Standard deviation	Minimum	Maximum	Total
Affective bursts stimuli									
Anger	354 Hz	101 Hz	196 Hz	514 Hz	71.4 dB	13.4 dB	36 dB	86 dB	1.02 s
Happiness	460 Hz	101 Hz	240.5 Hz	598 Hz	68 dB	9.6 dB	40.3 dB	82.4 dB	1.07 s
Neutral	217 Hz	21.6 Hz	207 Hz	379 Hz	80 dB	4 dB	54.2 dB	82.5 dB	1.56 s
Meaningless speech stimuli									
Anger 1	268 Hz	41.3 Hz	134.5 Hz	358.2 Hz	72 dB	7.2 dB	51.2 dB	83 dB	1.81 s
Anger 2	356 Hz	91.2 Hz	181 Hz	555.5 Hz	72 dB	7.3 dB	50.6 dB	82 dB	1.2 s
Anger 3	444 Hz	127.5 Hz	145.6 Hz	588.6 Hz	71 dB	8.8 dB	47.3 dB	83.5 dB	1.76 s
Happiness 1	450.5 Hz	122.5 Hz	148.3 Hz	588 Hz	70 dB	13.2 dB	13.2 dB	84.4 dB	1.77 s
Happiness 2	268 Hz	41.4 Hz	133 Hz	358 Hz	72 dB	7.2 dB	51.3 dB	83 dB	1.8 s
Happiness 3	358 Hz	92.5 Hz	181 Hz	556 Hz	72.5 dB	6.7 dB	56 dB	82 dB	1.2 s
Neutral 1	207.5 Hz	44.5 Hz	145.5 Hz	333 Hz	70.7 dB	7.5 dB	54 dB	84 dB	2.1 s
Neutral 2	191.3 Hz	32.3 Hz	146.6 Hz	245 Hz	71 dB	7.8 dB	50 dB	85.8 dB	2.3 s
Neutral 3	155.5 Hz	10.4 Hz	136.3 Hz	189.4 Hz	71 dB	9 dB	29.6 dB	81.7 dB	1.75 s

The main effect of age, $F(3, 76) = 2.45$, $p = .069$, $\eta_p^2 = .088$, was not significant. A main effect of voice condition, $F(1, 76) = 7.70$, $p = .007$, $\eta_p^2 = .09$, was revealed: There were greater looking times at AOIs after no voice (1196 ± 40 ms) than after the neutral voice (1098 ± 43 ms). The main effect of emotional face was not significant, $F(1, 76) = 1.54$, $p = .219$, $\eta_p^2 = .020$. The main effect of AOIs, was not significant, $F(1, 76) = 0.08$, $p = .780$, $\eta_p^2 = .001$, and no significant difference of looking time at the eyes (1161 ± 65 ms) or the mouth (1132 ± 62 ms) was found.

However, a significant interaction between emotional faces and AOIs, $F(1, 76) = 16.53$, $p < .001$, $\eta_p^2 = .179$, was found. Post-hoc analyses using Bonferroni indicated that the happy mouth was looked at less than the happy eyes ($p = .008$) or the angry mouth ($p < .001$). We observed a significant interaction between age and emotional faces, $F(3, 76) = 3.20$, $p = .028$, $\eta_p^2 = .112$, and post-hoc analyses using Bonferroni indicated that the 5-year-olds looked less at the happy AOIs than did the adults ($p = .02$).

Experimental condition: Happy and angry voices.

Looking times at faces. We analyzed the results of the experimental condition for the looking times at faces presented after a happy or angry voice as a function of age group. A repeated-measures ANOVA was performed on looking times at faces with emotional faces (happy, angry) and voice condition (happy, angry) as within-subject factors and age (5, 8, 10 years and adults) as a between-subject factor.

A main effect of age, $F(3, 76) = 8.89$, $p = .00004$, $\eta_p^2 = .259$, was revealed. Post-hoc analyses using Bonferroni indicated that the 5-year-olds (3118 ± 167 ms) looked for a shorter time than the 8-year-olds (3773 ± 167 ms; $p = .04$), 10-year-olds (4015 ± 167 ms; $p = .002$) and adults (4280 ± 167 ms; $p < .001$). The main effect of voice condition was not significant, $F(1, 76) = 0.55$, $p = .461$, $\eta_p^2 = .007$. The main effect of emotional face was not significant, $F(1, 76) = 1.49$, $p = .226$, $\eta_p^2 = .019$.

However, the interaction between the voice condition and emotional faces was significant, $F(1, 76) = 33.88$, $p < .001$, $\eta_p^2 = .308$. A planned comparison suggested that after the happy voice, the happy face was looked at longer than the angry face ($p < .001$), and after the angry voice, the angry face was looked at longer than the happy face, ($p < .001$). Moreover, interaction among age, voice, and emotional faces was significant, $F(3, 76) = 2.74$, $p = .048$, $\eta_p^2 = .098$. A planned comparison suggested that after the happy voice, the happy face is looked at longer than the angry face in the 8-year-olds ($p = .004$), 10-year-olds ($p = .05$) and adults ($p < .001$), but no difference in looking time was found in the 5-year-olds ($p = .601$). After the angry voice, the angry face was looked at longer than the happy face in the 10-year-olds ($p < .001$) and adults ($p < .001$); in the 8-year-olds this preference in looking time tended to be significant ($p = .055$), but no difference in looking time was found in the 5-year-olds ($p = .389$).

Looking time at AOIs. We analyzed the results of the experimental condition for the looking time at AOIs (eyes, mouth) presented after the happy or angry voice as a function of age group. A repeated-measures ANOVA was performed on looking time with emotional faces (happy, angry), AOIs (eyes, mouth), and voice condition (happy, angry) as within-subject factors and age (age 5, 8, 10 and adults) as between-subjects factor.

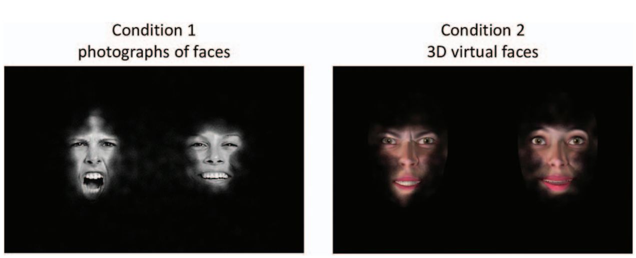

Figure A1. Focus map: all participants' gaze patterns at the stimulus image visualized as a transparent map. Visual stimulus Condition 1 with copyright persmission from Lundqvist, Flykt, and Öhman (1998); visual stimulus of Condition 2 with permission from Roesch et al. (2011). See the online article for the color version of this figure.

(*Appendix continues*)

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.

A main effect of age, $F(3, 76) = 3.49$, $p = .02$, $\eta_p^2 = .121$, was found. Post-hoc analyses using Bonferroni indicated that the adults (1200 ± 75 ms) looked longer at the AOIs than did the 5-year-olds (874 ± 75 ms; $p = .019$). The main effect of voice condition, $F(1, 76) = 1.66$, $p = .202$, $\eta_p^2 = .021$, was not significant. The main effect of emotional face was not significant, $F(1, 76) = 0.25$, $p = .616$, $\eta_p^2 = .003$. The main effect of AOIs, $F(1, 76) = 0.31$, $p = .578$, $\eta_p^2 = .004$, was not significant; the looking time at eyes (1096 ± 76 ms) was not significantly different from that of the mouth (1031 ± 64 ms).

A significant interaction between emotional faces and AOIs, $F(3, 76) = 4.76$, $p = .032$, $\eta_p^2 = .059$, was revealed. However, post-hoc analyses using Bonferroni indicated no significant differences of looking time between the happy or angry mouth and eyes, all *p*s >.05. A significant interaction between voice condition and emotional faces, $F(1, 76) = 29.75$, $p < .001$, $\eta_p^2 = .281$, was found. A planned comparison indicated that after the happy voice, the AOIs of the happy face were looked at longer than the AOIs of the angry face ($p < .001$) and that after the angry voice, the AOIs of the angry face were looked at longer than the AOIs of the happy face ($p < .001$). Moreover, a significant interaction among age, emotional faces, and AOIs, $F(3, 76) = 3.41$, $p = .021$, $\eta_p^2 = .119$, was revealed. Post-hoc analyses using Bonferroni suggested that the 5-year-olds looked less at the eyes than the 8-year-olds and adults looked at the happy eyes; the 8-year-olds looked longer at the happy eyes than at the mouth, all *p*s <.05.

Looking Time Analysis of Condition 2: Meaningless-Speech Voices and 3D Virtual Faces

Baseline condition: Looking time at AOIs after no voice and neutral voice. We analyzed the results of the baseline condition for the looking times at AOIs (eyes, mouth) presented after the no-voice or neutral voice condition as a function of age group. A repeated-measures ANOVA was performed on looking time with emotional faces (happy, angry), AOIs (eyes, mouth), and voice condition (no voice, neutral voice) as within-subject factors and age (5, 8, 10 years and adults) as a between-subjects factor.

The main effect of age, $F(3, 76) = 4.44$, $p = .006$, $\eta_p^2 = .149$, was significant. Post-hoc analyses using Bonferroni indicated that the adults (1483 ± 64 ms) looked longer at AOIs than the 5-year-olds (1158 ± 64 ms; $p = .004$). A main effect of voice condition, $F(1, 76) = 8.912$, $p = .004$, $\eta_p^2 = .105$, was revealed: There was greater looking time at AOIs after the no voice condition (1363 ± 36 ms) than after the neutral voice condition (1251 ± 39 ms). The main effect of emotional face was also significant, $F(1, 76) = 9.46$, $p = .003$, $\eta_p^2 = .111$; happy AOIs (1419 ± 53 ms) are looked at longer than angry AOIs (1194 ± 44 ms). The main effect of AOIs, $F(1, 76) = 23.50$, $p < .001$, $\eta_p^2 = .236$, was significant, suggesting a greater looking time at the eyes (1582 ± 77 ms) than the mouth (1032 ± 50 ms).

We observed a significant interaction between age and voice condition, $F(3, 76) = 3.69$, $p = .015$, $\eta_p^2 = .127$: Post-hoc analyses using Bonferroni indicated that the 5-year-olds looked less at the AOIs after the neutral voice than the 8-year-olds after no voice ($p = .04$) and the adults after no voice ($p = .002$) and neutral voice ($p < .001$). After the neutral voice, the 8-year-olds looked less at the AOIs than the adults ($p = .04$). We observe a significant interaction between age and emotional faces, $F(3, 76) = 4.03$, $p = .010$, $\eta_p^2 = .137$; post-hoc analyses using Bonferroni indicated that the adults looked longer at the happy AOIs than did the 5-year-old, ($p < .001$). Also, the adults looked longer at the happy AOIs than the 5-year-olds ($p = .004$), 8-year-olds ($p < .001$), 10-year-olds ($p = .008$), and adults ($p = .008$) looked at the angry AOIs. A significant interaction among age, emotional faces, and AOIs, $F(3, 76) = 4.34$, $p = .007$, $\eta_p^2 = .146$, was revealed. Post-hoc analyses using Bonferroni indicated that the 8-year-olds looked longer at the happy eyes than the happy or the angry mouth; the 10-year-olds looked longer at the happy and angry eyes than at the angry mouth; and the adults looked longer at the happy eyes than the angry eyes, happy mouth, or angry mouth.

Experimental condition: Happy and angry voices.

Looking times at faces. We analyzed the results of the experimental condition for the looking times at faces presented after the happy or angry voice as a function of age group. A repeated-measures ANOVA was performed on looking times at faces with emotional faces (happy, angry) and voice condition (happy, angry) as within-subject factors and age (5, 8, 10 years and adults) as a between-subjects factor.

A main effect of age, $F(3, 76) = 6.11$, $p < .001$, $\eta_p^2 = .194$, was revealed. Post-hoc analyses using Bonferroni indicated that the adults (4465 ± 165 ms) looked longer than the 5-year-olds (3501 ± 165 ms; $p < .001$) and 8-year-olds (3766 ± 165 ms; $p = .02$). The main effect of voice condition was not significant, $F(1, 76) = 1.18$, $p = .282$, $\eta_p^2 = .015$. The main effect of emotional face was not significant, $F(1, 76) = 0.61$, $p = .439$, $\eta_p^2 = .008$.

However, the interaction between voice condition and emotional faces was significant, $F(1, 76) = 30.54$, $p < .001$, $\eta_p^2 = .286$. A planned comparison suggested that after the happy voice, the happy face was looked at longer than the angry face ($p < .001$) and after the angry voice, the angry face was looked at longer than the happy face, ($p < .001$). Moreover, the interaction among age, voice, and emotional faces was significant, $F(3, 76) = 8.09$, $p < .001$, $\eta_p^2 = .242$. A planned comparison suggested that after the happy voice, the happy face was looked at longer than the angry face in the 8-year-olds ($p = .003$), 10-year-olds ($p = .008$), and adults ($p < .001$), but no difference of looking time was found in the 5-year-olds ($p = .985$). After the angry voice, the angry face was looked at longer than the happy face in the 10-year-olds ($p = .01$) and adults ($p < .001$), but no difference in looking time was found in the 5-year-olds ($p = .795$) or 8-year-olds ($p = .755$).

Looking times at AOIs. We analyzed the results of the experimental condition for the looking times at AOIs (eyes, mouth) presented after the happy or angry voice, as a function of age group. A repeated-measures ANOVA was performed on looking times with emotional faces (happy, angry), AOIs (eyes, mouth), and voice condition (happy, angry) as within-subject factors and age (5, 8, 10 years and adults) as a between-subjects factor.

(*Appendix continues*)

A main effect of age, $F(3, 76) = 3.95$, $p = .011$, $\eta_p^2 = .135$, was found. Post-hoc analyses using Bonferroni indicated that the adults (1364 ± 71 ms) looked longer than the 5-year-olds (1061 ± 71 ms; $p = .02$) and 8-year-olds (1073 ± 71 ms; $p = .03$). The main effect of voice condition, $F(1, 76) = 0.35$, $p = .556$, $\eta_p^2 = .005$, was not significant. The main effect of emotional face was not significant, $F(1, 76) = 0.51$, $p = .479$, $\eta_p^2 = .007$. The main effect of AOIs, $F(1, 76) = 21.22$, $p < .001$, $\eta_p^2 = .218$, was significant: There was more looking time at the eyes (1399 ± 77 ms) than the mouth (909 ± 48 ms).

However, a significant interaction between voice condition and emotional faces, $F(1, 76) = 32.22$, $p < .001$, $\eta_p^2 = .298$, has been found. A planned comparison indicated that after the happy voice, the AOIs of the happy face were looked at longer than the AOIs of the angry face ($p = .001$), and after the angry voice, the AOIs of the angry face were looked at longer than the AOIs of the happy face ($p < .001$). Moreover, a significant interaction among age, emotional faces, and voices, $F(3, 76) = 9.55$, $p < .001$, $\eta_p^2 = .274$, was revealed. A planned comparison suggested that after the happy voice, the happy AOIs were looked at longer than the angry AOIs in the 8-year-olds ($p = .007$), 10-year-olds ($p = .058$), and adults ($p < .001$), but no difference of looking time was found in the 5-year-olds ($p = .542$). After the angry voice, the angry AOIs were looked at longer than the happy AOIs in the 10-year-olds ($p = .01$) and adults ($p < .001$), but no difference of looking time was found in the 5-year-olds ($p = .617$) or 8-year-olds ($p = .513$).

Received June 12, 2019
Revision received February 14, 2020
Accepted February 16, 2020 ■

This document is copyrighted by the American Psychological Association or one of its allied publishers.
This article is intended solely for the personal use of the individual user and is not to be disseminated broadly.