



HAL
open science

Infant sensitivity to mismatches between same/other-race faces and native/non-native speech

Anna Krasotkina, Julia Dillmann, Michael Vesker, Olivier Clerc, Olivier
Pascalis, Gudrun Schwarzer

► **To cite this version:**

Anna Krasotkina, Julia Dillmann, Michael Vesker, Olivier Clerc, Olivier Pascalis, et al.. Infant sensitivity to mismatches between same/other-race faces and native/non-native speech. *Infant Behavior and Development*, 2024, 77, pp.101997. 10.1016/j.infbeh.2024.101997 . hal-04722557

HAL Id: hal-04722557

<https://hal.science/hal-04722557v1>

Submitted on 9 Oct 2024

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

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

Infant sensitivity to mismatches between same/other-race faces and native/non-native speech.

Krasotkina A, Dillmann J, Vesker M, Clerc O, Pascalis O, Schwarzer G.

Infant Behav Dev. 2024 Sep 28;77:101997. doi: 10.1016/j.infbeh.2024.101997.

Abstract

Perceptual narrowing typically occurs around 6 months of age, and drastically changes an infant's perception of stimuli such as faces or spoken language according to the frequency with which the infant encounters them. It has already been well established that perceptual narrowing improves the sensitivity of infants to frequently encountered stimuli such as same-race faces and their native language while reducing their sensitivity to other-race faces and non-native languages. However, the effect of perceptual narrowing on the combined perception of face and language stimuli is not well understood. Therefore, to investigate the changes in the sensitivity of infants to matches and mismatches between faces and speech which might occur in the course of perceptual narrowing, we tested 3- and 9-month-old German infants using German faces and German spoken sentences which would be more familiar to the infants, as well as completely unfamiliar Chinese faces and French spoken sentences. The infants were tested using an intermodal association paradigm, whereby each infant saw sequences of German or Chinese faces, interspersed with German or French spoken sentences. We analyzed the total looking time of infants in conditions where the faces and speech were congruent (either both familiar, or both unfamiliar), versus incongruent conditions where only the faces or only the sentences were familiar. We found that while the 9-month-olds looked for similar durations in congruent versus incongruent conditions, the 3-month-olds looked significantly longer during congruent conditions versus incongruent conditions, indicating a greater attentiveness to face-speech matches and mismatches prior to the onset of perceptual narrowing.

Introduction

Infants are born with face- and speech-perception systems that are not yet attuned to a particular environment, and allow them to distinguish among both familiar types of stimuli which are common in that particular environment, and unfamiliar stimuli which are uncommon. However, during the first year of life, infants become more efficient at distinguishing among stimuli that they frequently encounter in their environment (Nelson, 2001), while losing the ability to discriminate among infrequently encountered stimuli. This process of attunement of infants' perceptual systems to stimuli according to their frequency of occurrence is known as perceptual narrowing.

In the face perception domain, perceptual narrowing occurs in infants who live in mono-ethnic environments between 3 and 9 months of age. At 3-months, infants typically do not yet exhibit differences in discrimination ability between same-race faces and faces from various other races (Kelly et al., 2007; Kelly et al., 2009). However, by the age of 9 months, the infants' face perception system becomes attuned to the faces they encounter most frequently, and thus 9-month-old infants are able to discriminate only between same-race faces (Kelly et al., 2007; Kelly et al., 2009; Krasotkina, Götz, Höhle, & Schwarzer, 2020; Krasotkina, Götz, Höhle, & Schwarzer, 2021). From this age onward, infants exhibit an "other-race effect": a diminished ability to identify persons from ethnicities that one does not have sufficient exposure to (for a review, see Rossion & Michel, 2011).

Around the same time, the infants' speech perception system becomes similarly attuned to the various types of speech sounds that are frequently encountered in the infants' environment. Infants start to show a decrease in sensitivity for non-native speech elements such as vowels at around 6 months, (Polka & Werker, 1994), to consonants at around 10–12 months, (Werker & Tees, 1984), to word stress at 10 months (Bijeljac-Babic et al., 2012), and to lexical tone contrasts between 6 and 9 months (Götz et al., 2018; Yeung et al., 2013). Thus, the infant's initially broadly tuned speech-processing system gradually becomes attuned to the set of speech sounds that are common in their native language.

The similarity in the timeframe of perceptual narrowing in the areas of face and speech perception, and that development in both domains is likely driven by an urge to communicate (Pascalis et al., 2014), has led to hypotheses about the existence of shared underlying domain-general mechanisms, that might drive perceptual narrowing in these domains (Hadley et al., 2014; Pascalis et al., 2014; Scott et al., 2007). It was shown, that perceptual narrowing for other-race faces and non-native tones occurs in the same infants between 6 and 9 months of age (Krasotkina, Götz, Höhle, and Schwarzer 2018). Moreover, the authors showed that the infants' ability to distinguish between other-race faces and their ability to distinguish between non-native speech sounds were correlated: Nine-month-old Caucasian infants, who had already undergone perception narrowing for non-native language contrasts showed decreased discrimination abilities for the other-race faces. A similar effect was observed in distinct cohorts of 9- and 12-month-old Japanese infants, whose discrimination abilities for other-race faces and non-native speech were correlated (Xiao et al., 2018).

A question which springs to mind is what evolutionary benefit could such domain-general processes bring? One simple possibility is that the sharing of such mechanisms across the face- and speech-perception domains could save on the resources needed to develop and support the required neural circuitry. However, an additional possibility is that such domain-general mechanisms could facilitate perceptual narrowing for stimuli which are by their nature multisensory. For instance, processing speaking faces, requires one to process both the audible speech and the visual information relating to facial mimicry, as evidenced by the well-known McGurk effect, where the shape and movement of the mouth modifies the perception of speech-sounds (MacDonald & McGurk, 1978). This effect has been demonstrated even in prelinguistic infants as young as 4.5 months (Burnham & Dodd, 2004). It therefore stands to reason that the infants' face- and speech-perception systems must at some point be calibrated to the facial articulation that accompanies the specific speech-sounds used in their native language through perceptual narrowing, as described above.

Evidence for this sort of multisensory perceptual narrowing comes from infants' ability to perceive matches and mismatches between the visuals of speaking same- and other-race faces and the sounds of native and non-native speech. For instance, in a study by Kinzler et al. (2007), 5- to 6-month-old infants looked significantly longer at the silent face of a same-race actress, who previously spoke their native language compared to a non-native language. Similar results were shown by Kubicek et al. (2014) in 4.5-month-olds, who looked significantly longer at videos of a speaking same-race person, where the articulation matched

the spoken language compared to when the articulation was mismatched, whether or not that language was native to the infants. However, older 6-month-old infants in the same study only showed this matching preference when the person spoke in the infants' native language (Kubicek et al., 2014).

At the same time, infants have been shown to not only be able to match same-race faces with a native language, but also to match other-race faces with a non-native language. In a study by Uttley et al. (2013), six-month-old infants were shown same- and other-race faces side by side, while speech in either the infants' native language or a non-native language was played in background. Infants looked longer at the same-race faces when listening to speech in their native language, and vice-versa. Thus, infants at 6 months of age have some multisensory expectations of facial and linguistic cues.

Although studies such as those described above provide evidence on some of the dynamics of infants' perception abilities when it comes to multisensory matching of same- and other-race faces with native and non-native language, thus far there have been no studies which investigated the development of multisensory matching in infants before and after the onset of the perceptual narrowing process. Thus, the purpose of our study was to address this gap by testing infants at the age of 3 months, after which point perceptual narrowing has been reported to begin, and at the age of 9 months, when perceptual narrowing should already be clearly manifested. To facilitate comparison with previous research (Clerc et al., 2022; Krasotkina et al., 2018; Krasotkina et al., 2021 Kelly et al., 2007), we paired German speech (the native language of our participants) with French speech, and European German faces (the ethnic background of our participants) with Asian Chinese faces.

Using a between-subjects design, we showed 3-month-old and 9-month-old infants photographs of same-race or other-race faces, followed by sentences spoken in their native or non-native language. Each infant was presented with faces and sentences in one of 4 conditions which differ in term of familiarity in the pairing (2 congruents and 2 incongruents): same-race faces combined with native language (congruent), other-race faces combined with non-native language (congruent), same-race faces combined with non-native language (incongruent), or other-race faces combined with native language (incongruent).

Although at 3 months old our youngest age group should already be old enough to show looking preferences for same-race faces (see Maurer and Werker, 2013 and Quinn et al., 2019 for comprehensive reviews), we do not anticipate a significant overall looking time

preference for same-race faces in our study, because it is focused on contrasts between combinations of faces and speech rather than contrasting same- and other-race faces. Instead, we believe that at least some of the infants may show a preference for congruent combinations of faces and speech over incongruent combinations.

Specifically, we hypothesized that since 3-month-old infants have not yet undergone the perceptual narrowing process and thus did not yet form any multisensory expectations, they would show similar looking times in the congruent and incongruent conditions. On the other hand, we expected 9-month-olds, who should have already undergone perceptual narrowing, to look longer at congruent trials compared to incongruent trials.

Materials and Methods

Ethics statement

The present study was carried out in accordance with the Research Ethics Guidelines of the German Psychological Society (DGPs). The experimental procedure and informed consent protocol were approved by the Research Ethics Office of the University of Giessen. Prior to testing, informed consent was obtained from all parents in writing for their children's participation in the study.

Participants

We tested 108 healthy full-term Caucasian infants, 54 of whom were 3 months old ($M = 105$ days; range: 90-125 days, 26 female and 28 male) and the other 54 were 9 months old ($M = 294$ days; range: 274-321 days, 33 female and 21 male). According to parental surveys, all the infants in the study were monolingual and had no exposure to persons of Asian descent. 12 additional infants were excluded from the final sample because of fussiness ($N = 9$), or technical problems during testing ($N = 3$).

Stimuli

Speech stimuli

We used 10-second recordings of sentences spoken in the infants' native German and non-native French as language stimuli. To avoid habituation effects, 6 female native-speakers voiced 6 sentences in each language. The sentences were phrases that are commonly used when communicating with infants (e.g., "Hi baby! How are you? Did you sleep well...?") and

were spoken in an infant-directed manner (i.e., “baby talk”). Stimulus recordings were played at a volume of approximately 55 dB.

Face stimuli

For the face stimuli we used color photographs of six other-race Asian (Chinese origin) and six same-race Caucasian (German origin) women, on a white background. In each photograph, the model smiled and looked at the camera with the hair, neck, and shoulders being visible. Each photograph was presented in the middle of the screen, appearing as 12.5 cm (10.98° visual angle) wide and 16.5 cm (14.47° visual angle) tall.

Procedure

Once the parents were informed about the general aims of the study, they granted explicit permission for their child to participate in our informed consent documentation, without being made aware of the specific hypotheses being tested. The infants were seated on their caregivers' laps, positioned approximately 60 centimeters away from a display screen with a resolution of 1920 x 1080 pixels and a Tobii Pro X3-120 eye-tracker with a data-sampling rate of 120 Hz positioned underneath the display. To ensure accurate eye-tracking data, the parents were provided with completely opaque glasses and instructed to remain still throughout both the eye-tracker calibration and the experiment. We used a 5-point (1 point in each of the 4 corners, and the fifth in the center of the screen) calibration procedure. Calibration was considered successful only if all 5 points were accurately calibrated, and we attempted calibration up to a maximum of three times. Infants who could not be successfully calibrated after three attempts were excluded from the final sample.

Infants from each of the two age groups were randomly assigned to one of four experimental conditions (A, B, C, and D, see Table 1). Infants in Group A were exposed to same-race faces and heard sentences in their native language of German, which represented a congruent condition, since both types of stimuli would be familiar. Infants in Group B were exposed to other-race faces and heard sentences in a non-native language of French, which also represented a congruent condition since both types of stimuli were unfamiliar. Infants in Group C were shown same-race faces but heard sentences in a non-native language of French, which represented an incongruent condition, with only the faces being of a familiar type. Lastly, infants in Group D were shown other-race faces and heard sentences in their native language of German, which also represented an incongruent condition with only the speech stimuli being familiar.

Table 1*Description of study groups*

Group	A	B	C	D
Congruency	Congruent	Congruent	Incongruent	Incongruent
Faces	same-race	other-race	same-race	other-race
Language	native	non-native	non-native	native
3-month-olds	n=13	n=15	n=12	n=14
9-month-olds	n=11	n=12	n=16	n=15

We utilized an intermodal association paradigm for the current study (Xiao et al., 2018). For all groups, the procedure began with a face stimulus being shown for 10 s. Next, a speech stimulus was played for another 10 s, followed by another face stimulus for 10 s, and so on, alternating between face and speech stimuli for 10 s each until all 6 face and speech stimuli have been presented. To avoid possible order effects, the order of presentation of the individual photographs and sentences was randomized. Figure 1 illustrates the procedure. Stimulus presentation was conducted using E-Prime version 2.0 (Psychology Software Tools, Pittsburgh, PA, USA).

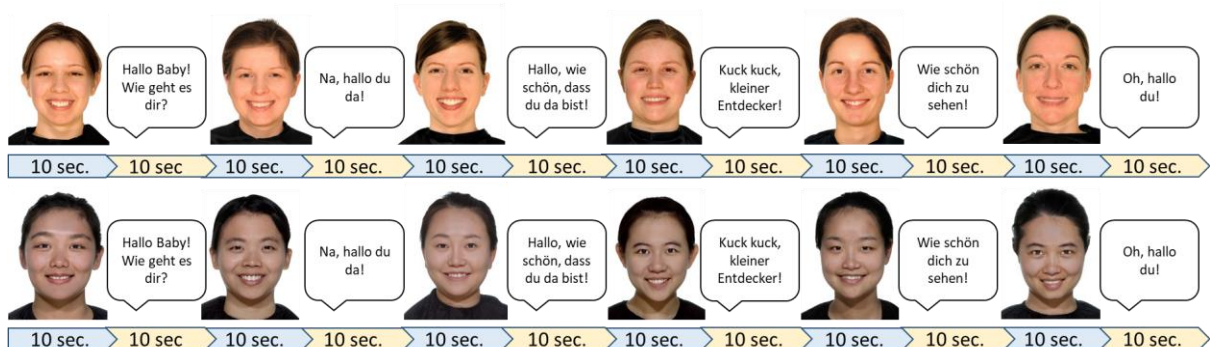


Figure 1 Possible stimulus presentation orders in the two conditions. In group A (top) same-race German faces were presented interspersed with native language German stimuli (congruent condition). In group D (bottom) other-race Chinese face were interspersed with native language German stimuli (incongruent condition).

Data analysis

The eye-tracking data were analyzed with Tobii Pro Studio using the included ClearView fixation filter. Based on previous analyses of infants' eye tracking data (Liu et al., 2011; Xiao, Quinn, Pascalis, & Lee, 2014), a fixation was defined as a minimum looking period of 100 ms within a radius of 30 pixels (0.765°). To analyze infants' gaze behavior, we calculated the total fixation time as the sum of the durations (in seconds) of all fixations made by each particular infant within an AOI covering the entire display for each trial.

Results

We began with a preliminary analysis testing for any baseline differences in looking time between same- and other-race faces prior to the infants hearing any of the speech stimuli. To do this we examined looking times during the first face stimulus in each trial sequence. We conducted a one-way ANOVA on looking time in sec. for the first face stimulus as the dependent variable with face stimulus type (same-race or other-race) and age group (3 months or 9 months) as independent variables. The analysis revealed only a significant main effect of age [$F(1,104) = 7.107, p = .009$], indicating that 3-month-old infants looked at the first face stimulus ($M = 5.221s, SD = 3.262$) significantly less than 9-month-olds ($M = 6.716s, SD = 2.506$). However, there was no significant effect of face stimulus type [$F(1,104) = 0.127, p = .722$] or interaction between face stimulus type and age group [$F(1,104) = 0.587, p = .445$], indicating that infants looked for similar durations at same-race ($M = 5.845, SD = 3.024$) and other-race ($M = 6.103, SD = 2.978$) faces prior to exposure to any speech stimuli.

Each infant's looking times for the subsequent five faces were added together to provide a sum of their total looking time in seconds for the five faces after speech stimuli were introduced. To examine the difference between infants' looking times we conducted a one-way ANOVA on the sum of infants' looking times in the 5 subsequent face trials in seconds, with age group (3 months or 9 months) face stimulus type (same-race or other-race) and speech stimulus type (native or non-native language) as between-subjects factors. We found a significant three-way interaction between age group, face stimulus type, and speech stimulus type [$F(1, 100) = 4.584, p = .035$], which indicated that 3-month-old and 9-month-old infants showed different looking

durations depending on the combination of face stimulus type and speech stimulus type (Figure 2). This analysis did not reveal any other significant main effects or interactions.

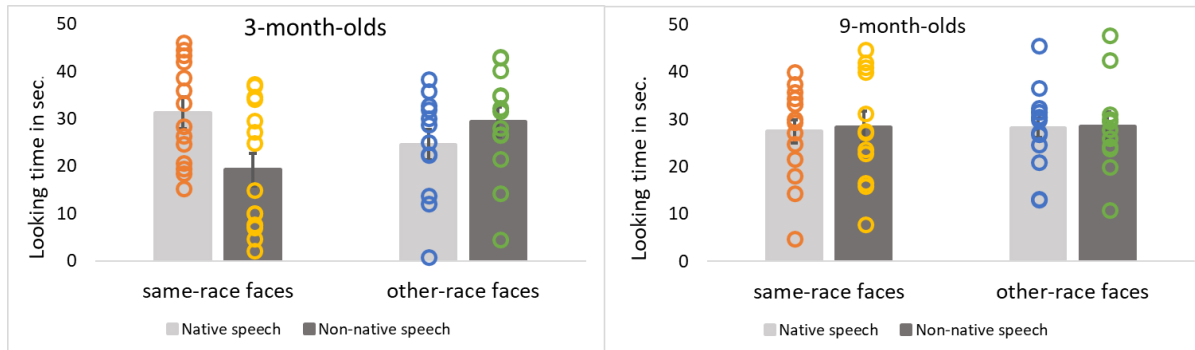


Figure 2 Infants' total looking times (in seconds) in two age groups (left: 3-month-olds, right: 9-month-olds). Each circle represents one infant's looking time. Error bars indicate standard error.

Examining Figure 2 above seems to suggest that what is driving the significance of the three-way interaction between age group, face type, and speech type, is that while the 9-month-old infants show similar total looking times for all conditions, the 3-month-olds seem to show longer total looking times in the congruent conditions versus the incongruent conditions. To confirm whether this was indeed the driver of the three-way interaction, we carried out a set of two post-hoc t-tests, comparing the infants' total looking time for the congruent versus incongruent conditions in each of the two age groups. The first t-test confirmed that the 3-month-old infants indeed looked significantly longer ($t(52)=2.779, p=.008$) at the face stimuli when they were combined with the speech stimuli congruently ($M=30.324, SD=10.751$), than when they were paired with the speech stimuli incongruently ($M=21.324, SD=12.319$). By contrast, the second t-test did not show any significant effect in the 9-month-old age group ($t(52)=-0.144, p=.886$): The 9-month-olds showed similar total looking times, regardless of whether the faces were combined with speech stimuli congruently ($M=27.912, SD=9.557$), or incongruently ($M=28.228, SD=9.600$) (Figure 3).

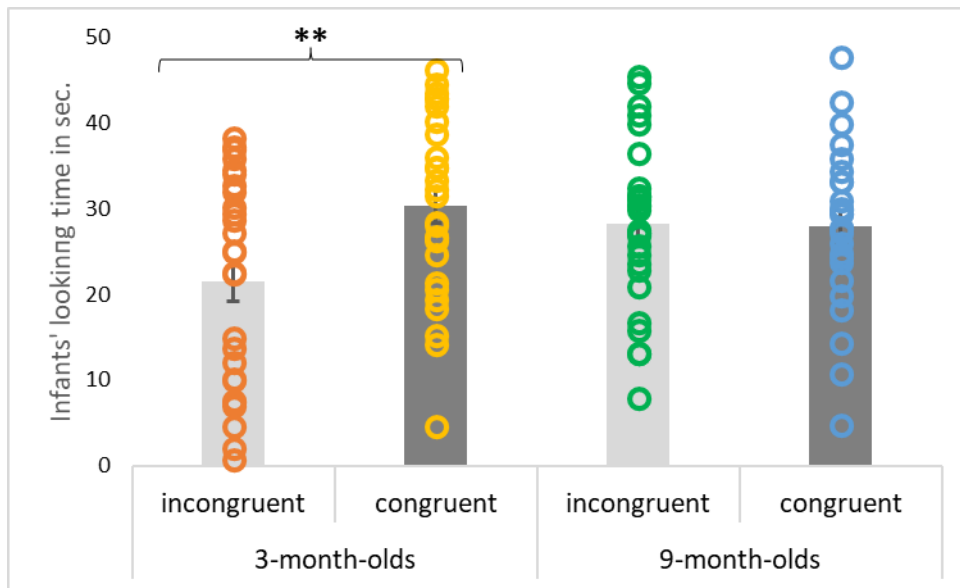


Figure 3 Infants' total looking times (in seconds) in two age groups (left: 3-month-olds, right: 9-month-olds) separated by condition (congruent or incongruent). Error bars indicate standard error, two stars indicate a significance level of $p < 0.01$.

Discussion

Our study used an intermodal association paradigm to examine how 3- and 9-month-old infants perceive intermodal associations between face and speech. We analyzed infants' looking times to investigate whether they had formed multisensory expectations of association between congruently matched same-race faces and native language, and between other-race faces and non-native language.

Our analysis regarding looking times for the first face stimulus seen by each infant regardless of the condition revealed only an effect of age, with the 9-month-olds looking at all faces significantly longer. We found no overall looking preference for either same-race or other-race faces, and no interactions between the age and face type. This was the result we had anticipated, since our experimental approach did not involve any simultaneous presentations of the two face types, or any familiarization or habituation stages, which are typically used to elicit the typical same-race looking preferences for infants established in the literature (see Maurer and Werker, 2013 and Quinn et al., 2019 for reviews).

However, the results of our main analysis were more surprising. We found that 3-month-old infants looked significantly longer during the congruent conditions than at the incongruent conditions. By contrast, the 9-month-old infants did not show any significant differences in looking time between congruent and incongruent combinations of faces and language stimuli.

These findings ran counter to our initial hypothesis, as we had initially supposed that the 3-month-old infants would not yet have any expectations regarding the combination of faces and languages, and would thus not show any significant difference in looking time between those combinations of stimuli. On the other hand, we supposed that by 9 months of age, infants would already be accustomed to the matching between same-race faces and their native language, and would therefore exhibit a different level of attention to congruent combinations of faces and speech versus incongruent combinations. However, our findings showed that it was in fact the 3-month-olds who were the most sensitive to differences between the congruent and incongruent combinations of faces and language, as evidenced by significant differences in looking time, while the 9-month-olds did not show such differences.

These findings resemble those of Pons et al. (2009) with infants from English and Spanish households, where the authors similarly found that younger infants showed a broader sensitivity to matches and mismatches between simultaneously presented videos of faces producing audible contrasting syllables. However, the younger infants in this study were 6 months of age, and would thus already have a significant amount of experience with perceiving speaking faces, allowing them to form some expectations regarding the facial movements associated with their native language. By contrast, the younger infants in our study were only 3 months of age, making it notable that they already form such expectations even at such a young age.

One possible explanation for these observations that the 3-month-olds were sensitive to congruency between faces and language, is that at this age they are in fact learning the correct facial mimicry associated with their native language, and are thus particularly sensitive to matches and mismatches between familiar faces types and their native language. Meanwhile, the 9-month-old infants would naturally be interested in congruent trials due to the potential for communication, but might also have exhibited an elevated level of attention to incongruent trials due to their sheer novelty.

This explanation for our observations could be investigated in future experiments by testing whether the sensitivity of infants to facial mimicry as it relates to spoken language does in fact change around the age of 3 months. The stimuli for such an experiment would need to disassociate the facial mimicry involved in spoken language from the other confounding factors present in our present study: Namely, the familiarity of the language and the ethnicity of the face. This could theoretically be accomplished by making use of recent advances in artificial

intelligence to create “deepfake” videos where the facial mimicry of a person of one ethnicity speaking a particular language is replaced by the mimicry of a person of another ethnicity speaking a different language, thereby allowing us to isolate the aspect of facial mimicry from both the rest of the face, and the language being spoken.

It is also important to note that our study focused on a particular combination of two languages (German and French) and two ethnicities (German and Chinese). Therefore, before we can confidently conclude that our findings (particularly our hypothesis regarding mimicry), reflect some fundamental aspect of an infants’ perceptual development, our results should first be replicated using stimuli from additional languages and ethnic groups.

Our results seem in contradiction with Xiao’s et al. study (2018 or 2017) where the intermodal association was between own or other race faces and music of different emotional valence (happy or sad). Music in major scale with fast tempo conveys happy emotional valence, whereas music in minor scale with slow tempo conveys sad emotional valence (Pesic, 2014). Infants from 5 months of age are able to discriminate the 2 category of music (Flom and Pick; 2012). They also can match happy music with happy facial expressions (Gentile, 1998; Nawrot, 2003). Taking advantage of this ability, Xiao et al. found no specific association between face race and music at 3 and 6-month of age but at 9 months infants look longer at own-race faces paired with happy music, and other-race faces paired with sad music. The difference in outcomes with our results might indicate that the association between face race and language emerges earlier than that between face race and music emotional valence. For language, infants might associate individual faces with the speech heard. In case of music, it could be that hearing happy music put them in an attentional mode that is different than sad music and it could be associated with the category seen and not the individual face.

In summary, our findings showed that after the typical onset period of perceptual narrowing, 9-month-old German infants showed similar looking times during both congruent and incongruent combinations of same/other-races faces and native/non-native speech stimuli. By contrast, prior to the usual onset period of perceptual narrowing, 3-month-old German infants looked significantly longer at congruent combinations of faces and language stimuli versus incongruent combinations, suggesting that this age group may be particularly sensitive to matches and mismatches between familiar and unfamiliar types of faces and speech.