Simulating newborn face perception

Olov von Hofsten
Eclipse Optics AB, Stockholm, Sweden

Claes von Hofsten
Department of Psychology, Uppsala University, Uppsala, Sweden

Unni Sulutvedt
Department of Psychology, University of Oslo, Oslo, Norway

Bruno Laeng
Department of Psychology, University of Oslo, Oslo, Norway

Tim Brennen
Department of Psychology, University of Oslo, Oslo, Norway

Svein Magnussen
Department of Psychology, University of Oslo, Oslo, Norway

A frequently asked question concerns what a newborn infant can actually see. The contrast sensitivity function of newborn infants is well known, but its implications for the ability of newborns to perceive faces of adults remain unclear. We filtered gray scale animations of facial expressions in terms of both spatial frequency and contrast to correspond to the properties of newborn infants’ acuity and showed them to adult participants. We reasoned that if adults were unable to identify the depicted facial expressions, then it would also seem unlikely that newborn infants could identify the same expressions. We found that for the simulated acuity the different expressions could be rather well identified at a distance of 30 cm, but when the distance was increased to 120 cm their discriminability was much degraded. This shows that although the perception of faces and facial expressions can function at the low visual resolution of the newborn infant, it is insufficient for distinguishing faces and facial expressions at moderate distances.

Introduction

A frequently asked question concerns what a newborn infant can actually see. Behavioral studies using the forced choice preferential looking technique (Teller, 1979; Teller, McDonald, Preston, Sebris, & Dobson, 1986), by which young infants’ visual acuity can be determined with great accuracy (Atkinson, Braddick, & Braddick, 1974; Banks & Salapatek, 1981; Brown & Yamamoto, 1986; Gwiazda, Brill, Mohindra, & Held, 1980), have shown that the spatial resolution of infants during the first few weeks is very crude. Spatial resolution is 1 to 2 cycles/deg of spatial frequency under optimal contrast conditions, which is around 2% to 3% of adult values. Figure 1 shows the contrast sensitivity of 4-week-old and 2- to 3-day-old infants compared with that of adults.

A number of graphic simulations have been produced to demonstrate what a newborn infant can see, but such demonstrations often fail in one or more respects. The simplest procedure is to arbitrarily blur the image, either by the use of a frequency threshold filter or by using a Gaussian blur, but this procedure takes into account neither the much lower contrast sensitivity of newborn children nor the resolution limit measured in controlled experiments, so the result cannot be close to a veridical representation of what an infant actually perceives. Contrast sensitivity considerations are more complicated and to our knowledge have not been treated previously in simulations of infant vision. Furthermore, it is of crucial importance that the resolution of the simulated image corresponds to what newborn infants actually see at a particular distance for a given field of view or a given object size. In addition, in the real world stimuli move and faces...
changes expressions. Thus, ecologically valid simulations need to take the factors of motion and change into consideration. Thus, the present paper is not concerned with the perceptual competencies of human infants but rather with the visual information available to newborn infants on which such competencies might be based, given the contrast sensitivity function (CSF) of newborns. We filtered dynamic images in terms of both spatial frequency and contrast for stimuli presented at different distances, selecting emotional facial expressions as stimuli. These stimuli were selected for two important reasons. First, face perception has been a major field of study in young infants (see, e.g., Farroni, Csibra, Simion, & Johnson, 2002; Johnson, Dziurawiec, Ellis, & Morton, 1991; Leo & Simion, 2009). Second, discriminating facial expressions represents a rather difficult everyday discrimination task. Small stimulus differences define different emotions. These differences might become undetected in stimuli filtered to the level of newborn acuity.

How might the world look to the newborn infant? It is evident that some structured vision is present at birth and that it guides certain behaviors. For instance, when an object is dangled in front of newborn infants, they direct their arm movements toward it (von Hofsten, 1982), and a wide angle flow field presented underneath newborn infants controls their stepping movements (Barbu-Roth et al., 2009). Face perception is somewhat more demanding and requires better visual resolution.

There is evidence that newborn infants perceive certain facial features because they track a moving schematic face more reliably than a scrambled face under otherwise equal conditions (Johnson, Dziurawiec, Ellis, & Morton, 1991; Goren, Sarty, & Wu, 1975). However, in these studies the contrast conditions were not controlled so that stimuli could have been presented at contrast levels that were better than those of a face under typical viewing conditions. Farroni et al. (2002) also showed that under optimal conditions, 1- to 5-day-old newborns prefer to look at faces looking straight back at them in comparison with faces with averted gaze. However, the question remains whether these results can be generalized to other viewing conditions.

In a much-debated article, Meltzoff and Moore (1977) claimed that newborn infants imitate tongue protrusion and mouth opening of an adult model. Many attempts have been made to replicate these effects but the results are mixed. Most positive effects have been obtained for tongue protrusion. In 21 out of 32 studies reviewed by Ray and Heyes (2011), significantly more tongue protrusions were elicited when the model performed this behavior. The attempts to replicate imitation of mouth opening were much less successful. Such an effect was found in only nine out of 29 studies (Ray & Heyes, 2011). The imitation studies have been discussed mostly in terms of whether an innate imitation mechanism is present or not. Another important consideration, however, is how well the newborn infant is able to perceive these expressions considering the poor spatial resolution and contrast sensitivity of infant vision found in studies of neonates. One factor that separates visual resolution and imitation experiments is the presence of motion or temporal change in the higher-order face perception experiments. It is well known from classical perception experiments that motion reduces perceptual ambiguity (Johansson, 1968). Recently, Skelton and Hay (2008) found that faces were recognized more accurately when presented in motion. In the present study, we therefore filtered animations of dynamically changing facial expressions rather than displaying still pictures.

We generated filtered images of faces showing different emotional expressions. The images were filtered with respect to both spatial frequency and contrast, in accordance with the CSF of 2- to 3-day-old infants (Atkinson, Braddick, and French, 1979; Brown & Yamamoto, 1986), and made to correspond to the resolution of images sensed by such an infant at different distances. Color perception has been found to be relatively undeveloped in newborns (Adams, Maurer, & Cashin, 1990; Hamer, Alexander, & Teller, 1982) and in order not to make any assumptions of the presence of color perception in newborns, gray scale photographs were used. We made animations of facial expressions that changed between a neutral expression.
and one of three emotional expressions—happy, surprised, and angry—and back to neutral again. As a control, we showed the morphing between the neutral expressions of the two models used. We presented these filtered animations to adult subjects and asked them what expressions they depicted. If they were able to see what the images depicted, we would conclude that it is possible that 2- to 3-day-old infants would be able to see it in a real face under identical viewing conditions—that is, if they are able to interpret those luminance changes in the same way as adults do. If the adult subjects cannot discriminate the facial expressions, it can be concluded that the infants cannot either.

If adults can identify these facial gestures, however, there is still a possibility that newborn infants cannot. In addition to the degradation of the stimuli caused by filtering, there are several possible differences between newborn infants’ ability to perceive faces and adults’ ability. There are, for instance, special mechanisms for the analysis of facial features that may not have matured in newborns (Hunnius, de Wit, Vrins, & von Hofsten, 2010). As we used dynamic stimuli, all the aspects of moving visual stimuli processed by adults may not be available to newborn infants (Wattam-Bell, 1991). Finally, there might be additional blurring caused by the suboptimally functioning visual accommodation. All these factors may eventually make visual perception of dynamic faces by newborn infants less correct than the estimation provided by our method of investigation. Thus, it is possible that the present method does not provide an accurate estimation of newborn acuity but rather a slight overestimation of it.

**Method**

**Subjects**

The study included a total of 48 participants (33 females) between the ages of 20 and 43 years (mean age = 24.18 years; SD = 5.96) with normal or corrected-to-normal vision. There were 16 subjects participating in each condition, and all were recruited at the Department of Psychology, University of Oslo, Norway.

**Stimuli and apparatus**

Fourier optics (see, e.g., Goodman, 2004) states that one can represent the transfer process of object to image by the modulation transfer function (MTF). It is a measure of the reduction in contrast from object to image over the spectrum, where contrast is defined as

\[
\text{contrast} = \frac{I_{\text{max}} - I_{\text{min}}}{I_{\text{max}} + I_{\text{min}}},
\]

where \(I_{\text{max}}\) and \(I_{\text{min}}\) are the maximum and minimum intensities of a sinusoidal function at a given frequency. Since all objects can be represented by a series of sinusoidal functions by use of the Fourier transform, the MTF reduces the amplitude of each such function of the object to produce the contrast distribution in an image. The MTF can be measured for an optical system, or can be theoretically calculated if relevant parameters are known.

The complete image formation calculation can be written as (Goodman, 2004)

\[
I(x, y) = \mathcal{F}^{-1} \left[ \text{MTF}(u, v) \mathcal{F} \left[ I_g(x, y) \right] \right],
\]

where \(\mathcal{F}\) denotes the Fourier transform, \(\mathcal{F}^{-1}\) is the inverse Fourier transform, \(I_g(x, y)\) is the unfiltered image intensity, and \(I(x, y)\) is the filtered image intensity defined in spatial coordinates \(x\) and \(y\). \(\text{MTF}(u, v)\) is the MTF, with \(u\) and \(v\) being the spatial frequency coordinates. It is clear that all information regarding the image formation process is contained in the MTF. It varies between zero, where all contrast is lost, and one, where no contrast is lost.

The image formation process in humans can be seen as a combination of these filters, where the aberrations and limitations of the eye’s optics, the detector or retina, and all image processing of the brain are included. Since the spatial filtering is linear, these filters can be combined into a single filter, the MTF of the entire system. With this function in hand, any image can be filtered to the same resolution and contrast as that viewed by a newborn infant. However, although the optics of the infant eye can be measured in theory and the retinal contribution to the MTF can be theoretically calculated, the newborn perception of the detected image is not available for analysis. What one can measure, however, is the CSF, as shown in Figure 1. It describes at what contrast a certain frequency grating becomes undetectable. Both the CSF and the system MTF share the same cutoff frequency, beyond which no contrast is transferred—the resolution limit of the system. In order to find the remaining MTF values so as to simulate image formation in infants, we make the following assumption:

\[
\frac{\text{MTF}_{\text{newborn}}}{\text{MTF}_{\text{adult}}} = \frac{\text{CSF}_{\text{newborn}}}{\text{CSF}_{\text{adult}}},
\]

where \(\text{MTF}_{\text{newborn}}\) and \(\text{MTF}_{\text{adult}}\) correspond to the contrast loss of the entire visual system. That is, we assume that the shift in contrast sensitivity between adults and infants is the same as the shift in contrast reduction. The MTF ratio for adults and infants can therefore be calculated from the known functions of the adult and infant contrast sensitivity. The result can
then be scaled with the MTF of the adult vision system to find $MTF_{\text{newborn}}$. The MTF ratio of Equation 3 for 2- to 3-day-old infants and 4-week-old infants is shown in Figure 2. At low frequencies (<0.1 cycle/deg), there are not enough empirical data for the calculation and we therefore set all values to one in the simulations.

To calculate the final image, the MTF ratio of Equation 3 is converted to a two-dimensional matrix, adapted for the desired distance, and used in Equation 2. Note that since we are using adult observers, the adult MTF is applied by the subjects themselves when observing the simulated images and is therefore not applied to the images in the simulation, although the difference is solely theoretical (the adult MTF is nearly unity for the considered spatial frequencies).

Four facial expressions (neutral, happy, angry, and surprised) shown in frontal, close-up gray scale photos of one man and one woman were selected out of the Karolinska Directed Emotional Faces (Lundqvist, Flykt, & Öhman, 1998). The original gray scale photographs were processed, using a Matlab script based on the outline above, to create images at different continuums of the spatial scale related to estimates of newborns’ visual acuity from three viewing distances (30, 60, and 120 cm). For each distance, the MTF ratio of adults to 2- to 3-day-old infants was applied to the original high-contrast images. The distances were selected so as to obtain images of low spatial frequency (1–6 cycles/image). The simulated images are shown in Figure 3 and were presented at a vertical and horizontal extension of 28.5 cm and 19.8 cm on the screen, approximately corresponding to the physical extension of a real face. The horizontal and vertical extensions were then 36.6° and 50.4° at a distance of 30 cm, 18.7° and 26.7° at a distance of 60 cm, and 9.4° and 13.5° at a distance of 120 cm.

We created six dynamic loops of the filtered happy, angry, and surprised facial images of both sexes using image-morphing software (Morpheus Photo Morpher, Morpheus Development, LLC, Ada, MI), moving smoothly from a neutral expression to an emotional expression and back to the neutral expression again. The Karolinska faces are particularly appropriate for this procedure since the position and size of the face are controlled and pixel changes occur locally. In addition, the high degree of filtration reduced the unnatural transitions that arise when using morphing software (Lander, Chuang, & Wickham, 2006). For control, we created an additional morphing from one of the neutral faces to the other neutral face and back again. Each animation lasted 2 s and was presented at the corresponding distances that they simulated. There are two independent factors that determine the transmission of information of the images: their resolution in terms of cycles per degree and the total area stimulated. At 30 cm the total area is 16 times larger than at 120 cm. The filtering takes care of the first factor and the presentation distance takes care of the second. The experiment was a $4 \times 2 \times 3$ design with four emotions, two sexes, and three distances. The order of the videos was pseudorandomized for practical reasons.

### Procedure

Participants were tested individually and informed about their rights as participants (Declaration of...
Helsinki. They were then seated in front of a computer screen at a distance of 30 cm, 60 cm, or 120 cm. Before the experiment, the participants were presented with a list of basic facial emotions: anger, fear, disgust, happiness, sadness, surprise, contempt, worry, arousal, and neutral. They were told that on each occasion, one of the two models used in the experiment was going to display an emotion. After each animation, the participant was given 6 s to write down the perceived emotion. Different groups of participants saw the stimuli at the three distances to avoid any priming. Thus, each participant saw eight emotions in a randomized order (four emotions by two models). Finally, the participants were debriefed and thanked for their participation.

Results

The results are shown in Figure 4 plotting the probability of a correct response for the different emotional expressions at the three simulated distances. Descriptive statistics were first computed for each condition and each participant, and all averages were analyzed by repeated-measures analysis of variance using StatView statistical software. The identification rate deteriorated substantially with simulated distance, $F(2, 45) = 49.78, p < 0.0001$. The emotional expressions were correctly identified in 75\%, 38\%, and 27\% of the trials at distances of 30, 60, and 120 cm, respectively. The simulated dynamic emotional expressions were identified to different degrees, $F(3, 135) = 5.68, p < 0.001$. The happy face was identified most often (i.e., on average in 61.5\% of the presentations). The other expressions were correctly identified in 41.6\%, 38.6\%, and 39.7\% of the trials at distances of 30, 60, and 120 cm, respectively. The simulated dynamic emotional expressions were identified to different degrees, $F(3, 135) = 5.68, p < 0.001$. The happy face was identified most often (i.e., on average in 61.5\% of the presentations). The other expressions were correctly identified in 41.6\%, 38.6\%, and 39.7\% of the trials for the angry, surprised, and neutral expressions, respectively. Finally, there was a strong interaction between distance and type of emotional expression, $F(6, 135) = 6.52, p < 0.0001$. This is shown in Figure 4. At the simulated distance of 30 cm, the happy face was correctly identified in 100\% of the presentations, the surprised expression was identified in around 75\% of the presentations, and the angry expression was identified in around 60\% of the presentations. The morphing...
between the neutral faces was identified in around 55% of the presentations. At 120 cm, the angry face was most often identified (56.3%), and the surprised expression was identified in only 6.3% of the presentations. This is shown in Figure 4. For all the expressions except for the angry one, the identification rate decreased substantially with simulated distance. For the angry expression, however, the identification rate was significantly higher for the 120-cm distance than for the 60-cm distance, \( t(15) = 3.873, p < 0.001 \).

Table 1 shows that the higher identification rate for the angry expression at 120 cm is not a question of a general preference for this expression. The other emotional expressions were not confused with the angry one; they were just more difficult to identify.

When averaging the responses over conditions, the male model and the female model were identified equally often (i.e., in 45.3% of the presentations). There were a couple of exceptions to this rule where the identification of the two models deviated more than 13% from each other. At 120 cm the happy face was identified in 100% of the trials at that distance. However, at the 120-cm simulated distance the identification rate was much lower. In fact, in most cases it was hardly greater than what is expected from random responding. An interesting exception is the identification of the angry expression. At this distance, it was identified in more than half the trials. This suggests that a special visual mechanism acting on very-low-resolution images might be engaged in this case. A study by Smith and Schyns (2009), where facial expressions were filtered across several bands of spatial frequency and then tested via the “bubbles” analysis.

### Discussion

Given that adults, when viewing the stimuli that simulated 2- to 3-day-old infants’ visual acuity, could rather well identify some of the expressions depicted at the 30-cm distance, the present results indicate that infants of that age have the possibility to perceive some dynamic facial expressions under optimal distance viewing conditions. For instance, the happy expression was judged to be sad in five cases at 120 cm. In a similar frequency filtered experiment by Du and Martinez (2011), happy and surprised were never (consistently) confused for other expressions regardless of the resolution. With our filtering, this was valid for the 30-cm simulated distance but not for the 60- and 120-cm distances. Surprise was incorrectly identified in 66% of the cases at the 60-cm simulated distance and in 98% of the cases at the 120-cm simulated distance.
method, revealed that at low spatial frequencies the expression of anger is conveyed by information restricted to the eye region whereas the expression of surprise is based on a narrow region corresponding to the mouth and that of happiness is based on viewing a combination of facial regions (eyes, cheeks, and mouth). It is then possible that, when the face image was highly degraded (at 120 cm), participants chose to monitor only the eye region, which would benefit the expression of anger but not that of happiness or surprise.

A remarkable fact is that faces can be identified even when the filtering is substantial. Sinha, Balas, Ostrovsky, and Russell (2006) found that subjects could identify more than half the set of familiar faces that had been blurred to have equivalent image resolutions of just $7 \times 10$ pixels, which is about the same resolution as the images depicting faces at 120 cm in the present study. However, when contrast is filtered in addition to spatial frequency as in the present study, the perception of faces degrades more rapidly.

We cannot, of course, from the present simulation experiments about infant perception, tell whether newborns have the necessary neural processing capability to make sense of the stimuli. The results of Farroni et al. (2002) and Leo and Simion (2009), however, indicate that very young infants have the capability to perceive complex facial properties. Leo and Simion (2009) showed that newborn infants recognize configural changes within real face images by testing their sensitivity to the Thatcher illusion. The faces were presented at a visual angle that corresponded to a presentation of a real face at a distance of 30 cm, as in one of the conditions in the present experiment. These results indicate that face perception might be conveyed by a special mechanism that is already functioning in newborn infants.

Another factor that might influence perception is optical defocus. We presented the faces at 30, 60, and 120 cm. Studies (Banks, 1980; Braddick, Atkinson, French, & Howland, 1979) show that infants younger than 1 month have accommodated for a viewing distance of 40 cm, with little capability of refocusing. Since the CSF data for the simulations were based on this distance, we assume an accommodated eye. There may therefore be a further decrease in resolution for a newborn observer at 120 cm due to focus inability.

There was a dramatic effect of viewing distance on the ability to perceive the facial expressions simulating 2- to 3-day-old infants’ acuity. At 120 cm, the resolution of the image was much impaired and the facial features were very hard to distinguish. The surprised expression did not show any signs of being detected at this simulated distance.

The facial expression that was easiest to identify was the smile. This suggests that humans may have a very robust mechanism both for displaying and for perceiving this positive facial gesture. Alternatively, this expression provides more salient features (e.g., there could be more pixel changes in a smile than in the other expressions we used). Within a few weeks from birth, infants both produce smiles and respond to smiles by smiling back (see, e.g., Darwin, 1872). The facial surprise gesture was the most difficult expression to identify with increasing distance. Even at the 60-cm simulated distance, it was identified in only one out of three cases. The most prominent feature of the facial surprise gesture was an open mouth. If the open mouth is difficult to perceive in the impoverished image, then this could explain why the studies of imitation of mouth opening have been so difficult to reproduce.

Using image filtering of spatial frequencies of stimuli to get insights into the efficacy of selected ranges of visual spatial information has been a fruitful approach to several different perception problems in addition to the present one (see Schyns & Oliva, 1999). Laeng et al. (2010), for instance, showed that presenting emotional expressions in only the lowest spatial frequencies (<6 cycles/image) while the rest of the image shows a neutral expression is sufficient to affect the social judgment of the portrayed person by adult participants even though the low-frequency information remains invisible to the observers. In a developmental study, Dobkins and Harms (2014) filtered upright and inverted facial images into high- and low-spatial-frequency images and showed them to 4- and 8-month-old infants and adults. They found that the inversion effect in the infants was significantly stronger for the high-spatial-frequency images than the low-spatial-frequency images. However, there was no such effect in adults.

The simulated images show the inferior contrast sensitivity of newborns relative to that of adults. In most cases, only high contrasts can be discriminated by newborns. Therefore, it is of utmost importance to specify the viewing conditions in order to evaluate how perceivable (i.e., in terms of luminance and contrast) a certain display is to a newborn child. We examined the earlier negative findings on neonatal imitation reviewed by Ray and Heyes (2011) and found that none of the 11 studies showing negative results and mentioned in their review had specified these key variables in a satisfactory way. It is therefore conceivable that the negative findings in at least some of those reports are due to poor viewing conditions.

### Conclusions

The CSF of infants who are a few days old demonstrates the limitations of their visual system...
compared with adult vision; these limitations have consequences for the visual information available to newborns for natural scenes. Our study gives an upper limit estimation of neonates’ visual resolution. If the adult subjects can perceive the facial expressions, newborn infants could, in principle, do it as well, but if the adults cannot perceive the facial expressions, the information is not there to be perceived. There may, however, be additional factors that further delimit neonates’ ability to discriminate facial expressions, such as limited processing capability for emotional expressions and limited motion perception.

The results demonstrate that at a close distance (30 cm), emotional facial expressions are rather well distinguished by an adult observer under conditions simulating neonatal acuity. However, when distance is increased, performance rapidly degrades. At a distance of 120 cm, facial expressions simulating newborn acuity are hardly perceivable to an adult. Adults, however, have a very advanced mechanism for perceiving faces (see, e.g., Atkinson & Adolphs, 2011), which is also efficient at low image resolutions (Harmon & Julesz, 1973); however, this mechanism is likely to develop over the first months of life. If face perception is less advanced in newborn infants, the ability to discriminate facial expressions should be more limited than found in the present study. The results of Acerra, Burnod, and de Schonen (2002), de Heering et al. (2008), Farroni et al. (2002), and Leo and Simion (2009), however, indicate that the mechanism for the analysis of facial expressions also functions for infants who are a few days old. We conclude that distance is a very severe constraint on neonatal visual perception and that this factor should be seriously considered in any study of neonatal perception.

Keywords: infant vision, face recognition, computational modeling

Acknowledgments

C. v. H., S. M., and U. S. were supported by grants from the Norwegian Directorate for Children, Youth and Family Affairs (06/34707) to the Cognitive Development Research Unit, Department of Psychology, University of Oslo. C. v. H. was also supported by EU StG12-312292-CACTUS (Gustaf Gredeböck) We thank Peter Unsbo, Linda Lundström, and Robert Rosén for their valuable discussions.

Commercial relationships: none.
Corresponding author: Olov von Hofsten.
Email: olov.vonhofsten@eclipseoptics.com.
Address: Eclipse Optics AB, Stockholm, Sweden.

References


