Stereoscopy Amplifies Emotions Elicited by Facial Expressions

Jussi Hakala and Jari Kätsyri
Department of Computer Science, Aalto University, Espoo, Finland

Jukka Häkkinen
Institute of Behavioural Sciences, University of Helsinki, Finland

Abstract
Mediated facial expressions do not elicit emotions as strongly as real-life facial expressions, possibly due to the low fidelity of pictorial presentations in typical mediation technologies. In the present study, we investigated the extent to which stereoscopy amplifies emotions elicited by images of neutral, angry, and happy facial expressions. The emotional self-reports of positive and negative valence (which were evaluated separately) and arousal of 40 participants were recorded. The magnitude of perceived depth in the stereoscopic images was manipulated by varying the camera base at 15, 40, 65, 90, and 115 mm. The analyses controlled for participants' gender, gender match, emotional empathy, and trait alexithymia. The results indicated that stereoscopy significantly amplified the negative valence and arousal elicited by angry expressions at the most natural (65 mm) camera base, whereas stereoscopy amplified the positive valence elicited by happy expressions in both the narrowed and most natural (15–65 mm) base conditions. Overall, the results indicate that stereoscopy amplifies the emotions elicited by mediated emotional facial expressions when the depth geometry is close to natural. The findings highlight the sensitivity of the visual system to depth and its effect on emotions.

Introduction

An observer is able to infer the mental state and intentions of a person based on facial expressions. For example, a smile might invite interaction, while a frown signals the opposite. Deciphering facial expressions is an aspect of mentalizing, or theory of mind, that is, the process of inferring the mental states of others. Emotional facial expressions have received substantial attention in a number of research fields due to their role in avoiding threat (e.g., Öhman & Mineka, 2001) and their importance in social interaction

Corresponding author:
Jussi Hakala, Department of Computer Science, Aalto University, Espoo, Finland.
Email: jussi.h.hakala@aalto.fi

(e.g., Frith, 2009) and evoking empathy (e.g., Carr, Iacoboni, Dubeau, Mazziotta, & Lenzi, 2003). In nonmediated real-life situations, the evolutionary benefits of mentalizing are clear because the other person is able to physically interact with the observer.

In mediated communication, however, the other person's mental state is less relevant to the observer's immediate behavior. Although a picture of a face might evoke the same cognitive representation as a live face, the observer is simultaneously aware of the medium, for example, aware that the face is presented as a picture on a computer display. Pönkänen, Alhoniemi, Leppänen, and Hietanen (2011) and Pönkänen et al. (2008) demonstrated that the human brain processes live faces and pictures of faces differently even at early stages of processing. Their first study measured the face-sensitive event-related potential component N170 elicited by faces and dummy faces in live and picture conditions. Faces elicited stronger N170 responses than dummy faces but only in the live condition. Their second study compared direct and averted gaze in live and picture conditions. Again, the elicited event-related potential difference between the direct and averted gaze was significant only in the live condition. Furthermore, in both studies, the live condition elicited differences in emotional valence self-ratings that were absent in the picture condition.

Photographs of facial expressions have been used in human emotion research for over a century (Watson, 2004), and static two-dimensional images of faces remain the method of choice in many emotion studies (cf. Ekman & Friesen, 1976; Goeleven, De Raedt, Leyman, & Verschuere, 2008). During the latter part of the 20th century, researchers began to focus increasing attention on other perceptual cues provided by facial expressions. In particular, the role of motion in recognizing emotional facial expressions has recently become a major area of interest (e.g., Krumhuber, Kappas, & Manstead, 2013). Because motion is omnipresent in most live situations, a dynamic presentation of a facial expression replicates the live situation more closely than a static presentation. One study found that dynamic presentation enhanced the perception of intense angry and sad facial expressions (Harwood, Hall, & Shinkfield, 1999), but the intense happy expressions in the experiment were so easy to recognize that no differences between the conditions were found. Ambadar, Schooler, and Conn (2005) demonstrated that motion facilitates the perception of all facial expressions (including happy expressions) in pictures in which the expressions of emotion are subtle. The authors attributed the benefit of motion to sensitivity to change rather than sensitivity to the characteristic movements associated with individual emotional expressions. Motion also significantly enhances the recognition of subtle emotions in animated facial expressions (Kätsyri & Sams, 2008). Later, research revealed that in addition to change sensitivity, the specific dynamics associated with facial expressions (i.e., motion cues) facilitate the recognition of emotions in pictures of facial expressions (Bould, Morris, & Wink, 2008). Because the motion cues improved recognition accuracy only for subtle facial expressions, the authors concluded that intense facial expressions are recognizable regardless of other cues. However, intense dynamic facial expressions elicit higher experienced emotional arousal (Sato & Yoshikawa, 2007) and activity in brain regions associated with emotional processing (Arsalidou, Morris, & Taylor, 2011) than do static facial expressions.

Similar to motion, the stereoscopic presentation of faces might modulate the interpretation of facial expressions because stereoscopic vision supports more accurate perception of the three-dimensional structure of a face. In both stereoscopic imaging systems and live situations, the visual system fuses the images from the two retinae of the observer into a single three-dimensional cyclopean percept. The internal representations of faces include three-dimensional information (Jiang, Blanz, & O'Toole, 2009), which might

explain why stereoscopy enhances facial recognition (Burke, Taubert, & Higman, 2007). Moreover, stereoscopy increases experienced presence when the depth magnitude is natural (IJsselsteijn, de Ridder, Hamberg, Bouwhuis, & Freeman, 1998; Takatalo, Kawai, Kaistinen, Nyman, & Häkkinen, 2011). The magnitude of depth in a stereoscopic image depends on the distance between the optical axes of the left and right cameras (i.e., the camera base). When the viewing distance and retinal size of the picture and the physical world scene are equal and the camera base is equal to the interpupillary distance of the observer, the capture and viewing geometries are identical. Increasing or decreasing the camera base results in the flattening or stretching of the image, respectively, along the depth axis. This distortion is typically characterized as the roundness factor (Devernay & Beardsley, 2010), that is, the ratio of the perceived depth from binocular disparity to actual object depth.

Surprisingly, the literature has neglected the role of stereoscopy in perceiving and responding to pictures of emotional facial expressions. In the present study, we investigate the connection between stereoscopy and the emotions elicited by emotional facial expressions. Our primary hypothesis is that stereoscopy amplifies the emotional response elicited by facial expressions due to the addition of three-dimensional cues and increased presence. In addition, we explore the effect of depth magnitude on the emotions elicited. Because the illusion of nonmediation is at a maximum at the natural depth magnitude, our secondary hypothesis is that the benefit of stereoscopy peaks at the natural camera base.

Method

Stimulus Acquisition

Stereoscopic photographs of two females (F1, F2) and two males (M1, M2) professional actors served as experimental stimuli. Figure 1 presents examples of the stimuli. We used two canon 5D Mark II (Canon Inc., Tokyo, Japan) cameras with a 50-mm f/1.4 USM lens attached in a parallel configuration to a beam splitter stereo rig. The distance from the right camera focal point to the eyes of the model was 80 cm. The actors were instructed to look straight at the right side camera lens at all times. The right side camera was positioned directly in front of the actors' face, while the left camera was positioned from 15 to 115 mm left of center at 25 mm intervals. Thus, the resulting camera bases were 15, 40, 65, 90, and
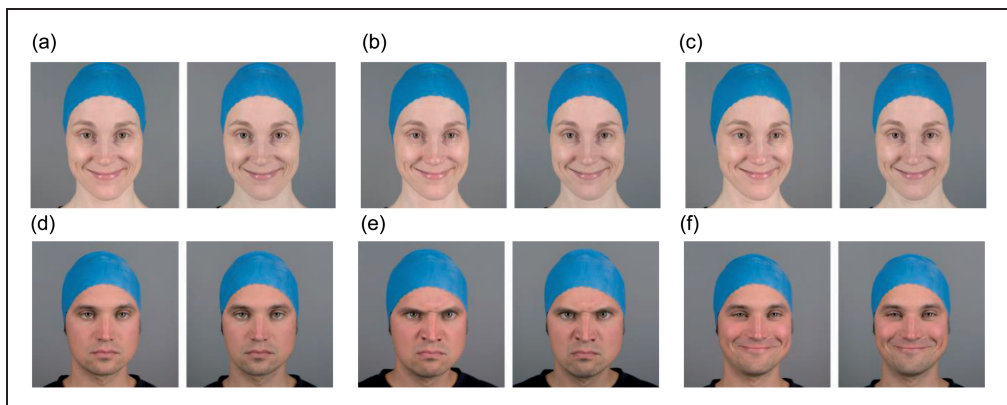

Figure 1. Sample stereoscopic stimuli arranged for parallel viewing. The top row exhibits a happy facial expression at the 15 mm (a), 65 mm (b), and 115 mm (c) camera bases. The bottom row exhibits neutral (d), angry (e), and happy (f) facial expressions at a 65 mm camera base.

115 mm. The actors wore a rubber cap to cover their hair and other external facial features that might otherwise introduce variability into the stimuli and draw attention away from internal facial features (Gronenschild, Smeets, Vuurman, van Boxtel, & Jolles, 2009). In addition to neutral facial expressions, we photographed angry and happy facial expressions. We also attempted to standardize the facial expressions across different actors and different shots of the same actor. Recording sessions were supervised by one of the authors (JK), who is a certified facial action coding system (Ekman, Friesen, & Hager, 2002) coder. The target facial configurations were $AU4 + 5 + 7 + 24$ (for anger) and $AU6 + 12$ (for happiness). The photographs were cropped to square aspect ratio and shifted, so that the midpoint between the actors' pupils in the left and right photographs was in the middle of the image. Thus, the eyes of the actors had zero stereoscopic disparity and were perceived at the display plane in the resulting stereoscopic image. The photographs were scaled to ensure that the distance between the pupils in the right side photograph on the display screen approximately matched the actors' interpupillary distance.

Stimuli and Apparatus

The experiment was conducted in a laboratory setting with controlled lighting. The display area surrounding the stimulus photograph and the wall behind the display were gray and had a luminance of $15 \, cd/m^2$. The display device was a 24-inch autostereoscopic display SL2400 (Tridelity AG, St. Georgen, Germany) positioned 80 cm in front of the participants. Figure 2 illustrates the effect of camera base on the perceived stimuli at the 80 cm viewing distance. To restrict participants' head movements and consequent image ghosting effects while viewing the autostereoscopic display, a chin rest was used throughout the experiment. Each participant was shown photographs of one male and one female actor. Participants were randomly assigned to one of the four different actor combinations (M1–F1, M1–F2, M2–F1, and M2–F2) such that an equal number of male and female participants saw each combination. The photographs were displayed under three stereo conditions: stereoscopic, monoscopic left, and monoscopic right. The three stereo conditions, five camera bases, three facial expressions, and two actors resulted in 90 trials per participant. Half of the participants viewed the original images, and the other half viewed images mirrored around the
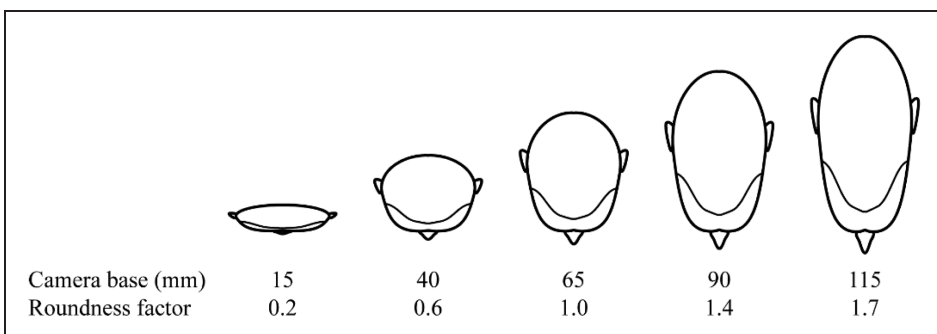

| Camera base (mm) | 15 | 40 | 65 | 90 | 115 |
| Roundness factor | 0.2 | 0.6 | 1.0 | 1.4 | 1.7 |

Figure 2. A simplified illustration of the effects of the camera base values used in the present study on the perceived depth cue from binocular disparity. The roundness factor, that is, the ratio of perceived depth from binocular disparity to the actual depth of the object, is calculated for an observer with a 65 mm interpupillary distance at 80 cm convergence distance. To observe the effect, see the 15, 65, and 115 mm stereoscopic sample stimuli presented in Figure 1.

vertical axis. The participants recorded their responses on a tablet computer that was attached to the table in front of them.

Participants

Twenty males and twenty females participated in the experiment; the mean age of participants was 28.1 years ($SD = 9.5$). The participants were of the same ethnic origin as the actors depicted in the stimuli. All participants exhibited normal or only slightly impaired visual acuity based on the Lea Numbers Test (Lea-Test Ltd, Helsinki, Finland). For four participants, visual acuity was below 20/20 (1.0 on the decimal scale) but above 20/40 (0.5). Near phoria was tested using the Maddox Wing test (Clement Clarke Ltd, London, UK); all participants measured less than 10 prism diopters horizontal and less than 2 prism diopters vertical phoria. All participants exhibited normal or only slightly impaired stereoacuity. Four participants failed the TNO stereo test (Laméris Ootech BV, Utrecht, The Netherlands) but performed well on the RANDOT stereo test (Stereo Optical Company Inc., Chicago, IL, USA). The mean interpupillary distance at the 80 cm viewing distance was 60.4 mm ($SD = 2.7$ mm). In addition to visual screening, participants completed the balanced emotional empathy scale (BEES; Mehrabian, 2000) and the 20-item Toronto alexithymia test (TAS-20; Bagby, Parker, & Taylor, 1994). The mean BEES score was 42.3 ($SD = 26.5$), and the mean TAS score was 45.3 ($SD = 9.2$). The present study adhered to the tenets of the World Medical Association Declaration of Helsinki and the ethical principles established by the Finnish Advisory Board on Research Integrity (http://www.tenk.fi/en/).

Procedure

Participants evaluated photographs that were displayed on the screen. No explicit time limit was imposed on the evaluations. Participants used a stationary tablet computer that was positioned in front of them to record their responses to the 9-point self-assessment manikin (SAM) valence and arousal scales (Bradley & Lang, 1994; Lang, 1980). They were instructed to evaluate the emotional experience evoked by the stimuli. Although the original SAM scale implicitly posits that positive and negative valence activation is reciprocal by measuring valence on a single bipolar scale, research has found that a single scale is sometimes insufficient because ambivalent emotions include both positive and negative valence components (Cacioppo, Gardner, & Berntson, 1997, 1999). Thus, in contrast to the original SAM scales, we separated the valence scale into positive and negative scales. Whereas the original scale ranged from unhappy to happy with intervening neutral manikins, the positive and negative scales used in the present study ranged, respectively, from neutral to happy and from neutral to unhappy manikins. No verbal labels were attached to the scale values. Furthermore, the arousal and valence figures were presented with the scale values increasing from left to right to facilitate the use of the scales. Other variables that were not within the scope of the present study were also measured. Participants completed two practice trials prior to the 90 randomized experiment trials. The mean duration of experimental sessions was 41 min ($SD = 9$ min); the stereoscopic trials lasted longer ($M = 28.6$ s, $SD = 11.3$ s) than the monoscopic trials ($M = 24.9$ s, $SD = 9.5$ s).

Analysis

The data were analyzed using R (R Core Team, 2015). To analyze the valence and arousal data, we used the lme4 R package (Bates, Maechler, Bolker, & Walker, 2015) implementation

of linear mixed models (LMMs) with restricted maximum likelihood estimates. LMM analysis was used rather than more conventional methods such as analysis of variance because it enabled us to more appropriately specify error variance, and LMM analysis can be viewed as a generalization of the more restricted analysis of variance (Quené & Van Den Bergh, 2004). The degrees of freedom were Satterthwaite approximations. The contrast between the stereo and mono conditions was calculated as the difference between the mean of the left and right monoscopic conditions and the stereoscopic condition. Significance tests of the differences between contrasts were performed using Holm–Bonferroni (Holm, 1979) adjusted p values. The predictors used in all the models were stereoscopy, camera base, facial expression, participants' gender, gender match (same/opposite), BEES score, and TAS score. The last four variables were used to control for potential confounds in the results because evaluations of emotional facial expressions have been found to be influenced by gender (Montagne, Kessels, Frigerio, De Haan, & Perrett, 2005), gender match (Vrana & Gross, 2004), empathy (Besel & Yuille, 2010), and alexithymia (Parker, Taylor, & Bagby, 1993). We also included interaction effects for all two-way interactions involving either stereoscopy or facial expression and all three-way interactions involving stereoscopy and facial expression. Random intercepts and facial expression slopes were defined for the individual actors to control for variability in individual actor expressions. In addition, random intercepts were specified for participants. The significance threshold was set to 5% in all analyses.

Results

Manipulation Check

To assess whether the facial expression stimuli elicited the expected emotional reactions, we analyzed the effects of facial expression type on negative valence, positive valence, and arousal ratings using LMM analyses. These manipulation checks were conducted for mean values that were pooled across monoscopic display conditions (i.e., left and right monoscopic images). Figure 3 presents comparisons of the marginal means of the emotional ratings of the facial expressions. Facial expression exerted a significant effect on negative valence, $F(2, 32) = 54.9$, $p < .001$, positive valence, $F(2, 16) = 108.4$, $p < .001$, and arousal, $F(2, 114) = 7.83$, $p < .001$. Contrasts between the facial expressions confirmed that the self-
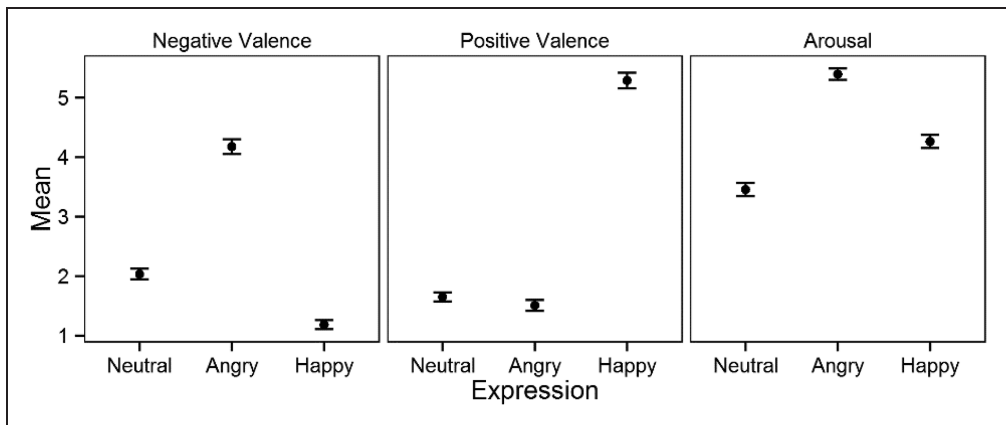

Figure 3. Marginal means of the three measured dimensions for the three facial expressions (95% CI).

reported emotions were consistent with the presented facial expressions. As expected, angry expressions elicited higher negative valence than neutral and happy expressions, and happy expressions elicited significantly higher positive valence than neutral and angry expressions. Neutral expressions also elicited higher negative valence than happy expressions. Angry expressions elicited higher arousal than happy and neutral expressions, and happy expressions elicited higher arousal than neutral expressions. We defined valence ambiguity, on a scale from 1 to 9, as the minimum of positive and negative valence ratings (Kaplan, 1972; see also Viinikainen, Kätsyri, & Sams, 2012). None of the stimuli elicited highly ambiguous emotions; the maximum ambiguity score was 2.15.

Negative Valence

The LMM analysis for negative valence revealed a significant three-way interaction between stereoscopy, camera base, and facial expression, $F(16, 3474) = 1.80$, $p = .030$. Figure 4 illustrates the results of planned contrasts between the stereoscopic and monoscopic conditions (averaged across the left and right image conditions) for different facial expressions at different camera bases. The stereoscopic neutral and happy expressions elicited significantly higher negative valence at the 90 mm and 115 mm camera bases compared with the monoscopic conditions. For the angry expression, stereoscopy significantly increased the negative valence at the 40, 65, and 90 mm camera bases but not at the extreme 15 mm and 115 mm camera bases.

We used planned comparisons to assess the extent to which differences in the camera bases influenced the effect of stereoscopic compared with monoscopic presentation. All comparisons were made to the 65 mm camera base, which was expected to provide the most natural depth percept because it was closest to the mean participant interpupillary distance. For the neutral expression, stereoscopy significantly increased the negative valence of all the other camera bases compared with the 65 mm camera base (Table 1). For the angry facial expression, stereoscopy increased the negative valence significantly more at the 65 mm camera base compared with the contrasts at the 15, 40, and 115 mm camera bases. The 65 mm contrast did not differ significantly from the contrasts at other
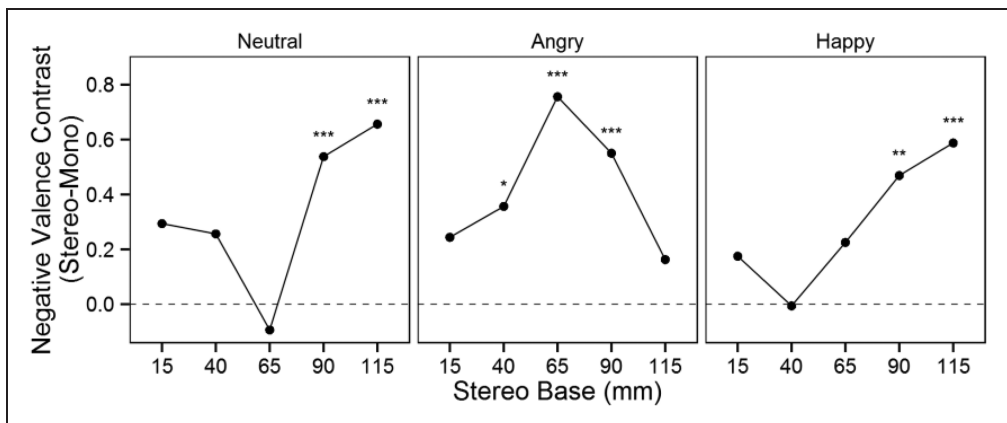

Figure 4. Negative valence contrasts between stereoscopic and monoscopic conditions.
*$p < .05$. **$p < .01$. ***$p < .001$, two-tailed.

Table 1. Contrast Comparisons for Negative Valence.

Expression	Contrast comparison	Estimate	SE	z ratio	Adj. p	Sig.
Neutral	15 vs. 65	0.387	0.160	2.417	.031	*
	40 vs. 65	0.350	0.160	2.183	.029	*
	90 vs. 65	0.631	0.160	3.937	<.001	***
	115 vs. 65	0.750	0.160	4.678	<.001	***
Angry	15 vs. 65	−0.512	0.160	−3.197	.004	**
	40 vs. 65	−0.400	0.160	−2.495	.025	*
	90 vs. 65	−0.206	0.160	−1.286	.198	
	115 vs. 65	−0.594	0.160	−3.704	<.001	***
Happy	15 vs. 65	−0.050	0.160	−0.312	.755	
	40 vs. 65	−0.231	0.160	−1.442	.298	
	90 vs. 65	0.244	0.160	1.520	.385	
	115 vs. 65	0.363	0.160	2.261	.095	

Note. The contrasts for each camera base were compared with the reference level for each facial expression; p values were Holm–Bonferroni adjusted to control for multiple comparisons for each expression.
*$p < .05$. **$p < .01$. ***$p < .001$, two-tailed.

camera bases for the happy expression, and stereoscopy exerted the smallest effect on negative valence at the 40 mm camera base.

Participant gender exhibited a significant interaction with facial expression, $F(2, 3377) = 8.5$, $p < .001$. Neutral and angry expressions elicited significantly higher negative valences in female participants compared with male participants. Gender match also exhibited a significant main effect, $F(1, 3474) = 19.6$, $p < .001$. Actors of the opposite gender elicited lower negative valences than actors of the same gender. Furthermore, facial expression exhibited significant two-way interactions with the BEES score, $F(2, 2848) = 12.2$, $p < .001$, and the TAS score, $F(2, 3476) = 10.1$, $p < .001$. Higher BEES scores predicted higher negative valences for neutral facial expressions, and higher TAS scores predicted lower negative valences for angry and neutral expressions. Although the main effects of camera base and facial expression were also significant, we interpret the effects of those predictors only within the significant interactions in which they were involved (see above).

Positive Valence

For positive valence, the three-way interaction between stereoscopy, camera base, and facial expression was not significant, $F(4, 3474) = 1.1$, $p = .377$. However, significant two-way interactions were found for stereoscopy and camera base, $F(8, 3473) = 2.6$, $p = .009$, and for camera base and facial expression, $F(8, 3474) = 2.0$, $p = .045$. Because the interaction between camera base and facial expression was inexplicable in itself and three-way interaction contrasts were planned, we analyzed the three-way interaction rather than the two-way interactions. Figure 5 presents the contrasts between the stereoscopic and monoscopic conditions. The neutral and angry expressions did not differ significantly between the stereoscopic and monoscopic conditions, but the contrast for the neutral facial expressions approached significance at the 115 mm camera base, $z = −1.87$, $p = .062$. For the happy expression, the 115 mm stereoscopic condition elicited significantly lower positive valence than the monoscopic condition, $z = −2.8$, $p = .005$. The contrasts for the 15, 40, and 65 mm camera bases in the happy expression also warranted further inspection. A post hoc contrast of the happy facial expression revealed that the combined
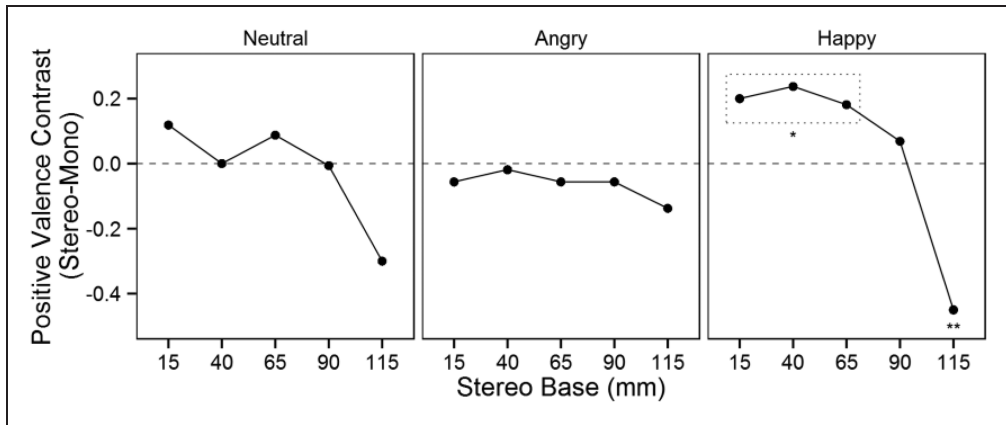
Figure 5. Positive valence contrasts between the stereoscopic and monoscopic conditions. The dotted rectangle highlights the three camera bases for the happy facial expression condition that, in combination, differed significantly between the monoscopic and stereoscopic conditions.
*$p < .05$. **$p < .01$, two-tailed.

Table 2. Contrast Comparisons for Positive Valence.

Expression	Contrast comparison	Estimate	SE	z ratio	Adj. p	Sig.
Neutral	15 vs. 65	0.031	0.161	0.194	.846	
	40 vs. 65	−0.088	0.161	−0.544	1	
	90 vs. 65	−0.094	0.161	−0.583	1	
	115 vs. 65	−0.387	0.161	−2.411	.064	
Angry	15 vs. 65	0.000	0.161	0.000	1	
	40 vs. 65	0.038	0.161	0.233	1	
	90 vs. 65	0.000	0.161	0.000	1	
	115 vs. 65	−0.081	0.161	−0.506	1	
Happy	15 vs. 65	0.019	0.161	0.117	.907	
	40 vs. 65	0.056	0.161	0.350	1	
	90 vs. 65	−0.113	0.161	−0.700	1	
	115 vs. 65	−0.631	0.161	−3.928	.000	***

Note. The contrasts for each camera base were compared with the reference level for each facial expression; *p* values were Holm–Bonferroni adjusted to account for multiple comparisons within each expression.
***$p < .001$, two-tailed.

positive valence of the three smallest camera bases (15, 40, 65 mm) was significantly higher in the stereoscopic condition than in the monoscopic condition, $z = 2.22$, $p = .026$. For the happy facial expression, stereoscopy increased the positive valence most at the 40 mm camera base. The planned comparison of contrasts using the 65 mm contrast as a reference yielded only a single significant difference following Holm–Bonferroni adjustment (Table 2); compared with the monoscopic condition, stereoscopy induced a significantly higher increase in positive valence for the happy expression at the 65 mm camera base than at the 115 mm camera base.

There were significant interactions between facial expression and BEES score, $F(2, 3438) = 11.2$, $p < .001$, and between facial expression and TAS score, $F(2, 3475) = 32.2$, $p < .001$. Higher BEES scores predicted higher positive valence ratings for neutral

expressions, whereas higher TAS scores predicted lower positive valence ratings for happy expressions. The interaction between gender match and facial expression was significant, $F(2, 3474) = 5.6$, $p = .004$. Angry expressions of the opposite gender elicited significantly higher positive valence compared with those of the same gender. The interaction between participants' gender and facial expression was also significant, $F(2, 3475) = 39.8$, $p < .001$. The happy facial expression condition elicited significantly higher positive valence in female participants compared with male participants.

Arousal

Figure 6 presents comparisons of the arousal contrasts between the stereoscopic and monoscopic conditions. Stereoscopy significantly increased arousal to the neutral facial expression for all but the smallest camera base. For angry facial expressions, stereoscopy increased arousal at the 65 mm and 90 mm camera bases. For happy expressions, the extreme camera bases (15 mm and 115 mm) in the stereoscopic condition elicited significantly higher arousal compared with the monoscopic condition. Comparison of the contrasts with the 65 mm reference contrast revealed no significant differences apart from the 115 mm camera base contrast for the happy facial expression condition (Table 3). However, a post hoc comparison of the 40 mm contrast to the other contrasts for the happy facial expression revealed significant differences between the 15 mm and 40 mm conditions, $z = 2.49$, $p = .038$, and between the 40 mm and 115 mm conditions, $z = 5.13$, $p < .001$. There was a significant interaction between facial expression and BEES score, $F(2, 3285) = 10.21$, $p < .001$. Higher BEES scores predicted lower arousal ratings for happy facial expressions. There was also a significant interaction between facial expression and TAS score, $F(2, 3477) = 6.16$, $p = .002$. Higher TAS scores predicted lower arousal ratings for neutral expressions and higher arousal ratings for happy facial expressions.

Discussion

We tested the hypotheses that stereoscopy enhances the emotional responses elicited by facial expressions, and that natural depth maximizes these effects. Stereoscopy significantly
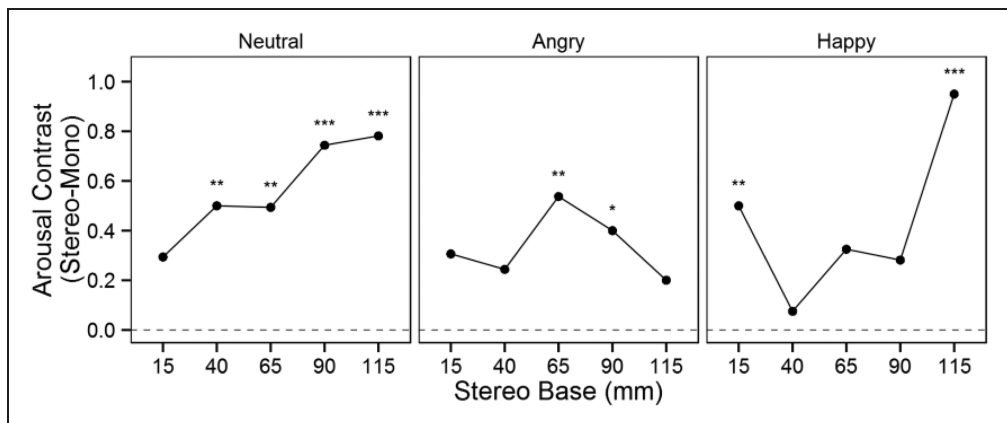

Figure 6. Arousal contrasts between the stereoscopic and monoscopic conditions.
*$p < .05$. **$p < .01$. ***$p < .001$, two-tailed.

increased the negative valence elicited by angry facial expressions, and the increase was highest for photographs taken at the 65 mm camera base. Similarly, stereoscopy increased the arousal elicited by angry facial expressions most at the 65 mm camera base. The results were less distinct for happy facial expressions. Stereoscopy subtly increased the positive valence of happy expressions in photographs taken at the 15, 40, and 65 mm camera bases compared with the monoscopic photographs. The combined contrast of the three camera bases was statistically significant between the stereoscopic and monoscopic conditions, but the individual contrasts were nonsignificant. In contrast to angry expressions, the arousal elicited by happy expressions was amplified by stereoscopy only at the extreme 15 and 115 mm camera bases.

Stereoscopy might enhance the emotional response of observers through two mechanisms. First, similar to motion, stereoscopy provides additional visual cues for the recognition of facial expressions. Earlier research has demonstrated that face recognition benefits from stereoscopic presentation (Burke et al., 2007), possibly due to three-dimensional information in internal representations (Jiang et al., 2009); this might also hold for facial expressions. The effect of stereoscopy in our results was clearest for angry facial expressions, which might indicate that the stereoscopic cues enhanced recognition of angry expressions, while happy expressions were relatively easier to identify without the additional cues provided by stereoscopy. This interpretation is supported by previous findings that demonstrated that happy facial expressions are unambiguously recognized, whereas angry facial expressions are more often confused with other negative emotions (e.g., Palermo & Coltheart, 2004). Our results further indicate that experienced valence and arousal are highly sensitive to the magnitude of depth; the benefit of stereoscopy reached a maximum at the natural depth magnitude. These findings suggest the operation of a second mechanism, which is the increased illusion of nonmediation, that is, presence. IJsselsteijn et al. (1998) found that stereoscopy increases presence if the observer perceives the depth as natural. In the case of stereoscopic facial expressions, both the physical and social components of presence (IJsselsteijn, de Ridder, Freeman, & Avons, 2000) are relevant. The appraisal of self-relevance is another process that modulates the perception of emotional facial expressions

Table 3. Contrast Comparisons for Arousal.

Expression	Contrast comparison	Estimate	SE	z ratio	Adj. p	Sig.
Neutral	15 vs. 65	−0.200	0.171	−1.172	.482	
	40 vs. 65	0.006	0.171	0.037	.971	
	90 vs. 65	0.250	0.171	1.465	.429	
	115 vs. 65	0.288	0.171	1.685	.368	
Angry	15 vs. 65	−0.231	0.171	−1.355	.351	
	40 vs. 65	−0.294	0.171	−1.721	.256	
	90 vs. 65	−0.138	0.171	−0.806	.420	
	115 vs. 65	−0.337	0.171	−1.978	.192	
Happy	15 vs. 65	0.175	0.171	1.026	.610	
	40 vs. 65	−0.250	0.171	−1.465	.429	
	90 vs. 65	−0.044	0.171	−0.256	.798	
	115 vs. 65	0.625	0.171	3.663	.001	**

Note. The contrasts for each camera base were compared with the reference level for each facial expression; p values were Holm–Bonferroni adjusted to account for multiple comparisons within each expression.
**$p < .01$, two-tailed.

(N'Diaye, Sander, & Vuilleumier, 2009). While social presence and self-relevance could be addressed separately (e.g., Blascovich et al., 2002), we propose that social presence increases self-relevance. As the physical and social presence elicited by a facial expression presented on a display screen increases, effectively transforming a picture into something more real, the stimulus becomes more self-relevant to the observer. Self-relevance might also explain the significantly stronger amplification of the valence elicited by stereoscopic angry facial expressions compared with happy facial expressions because angry expressions are more relevant to the immediate behavior of the observer than are happy expressions (Cacioppo et al., 1999).

The stereoscopic characteristics of the media present a challenge when assessing the effects of stereoscopy on emotions or presence. First, several artefacts (Boev, Hollosi, Gotchev, & Egiazarian, 2009) appear only in the stereoscopic rendition of the stimuli. These artefacts might be due to content creation, processing, or presentation. Second, the combination of the capture and viewing geometries affects the resulting depth percept. Although the roundness factor (Devernay & Beardsley, 2010) provides an estimate of the distortion due to binocular disparity, the visual system employs other depth cues, such as size, occlusion, and accommodation, that interact with each other. Individuals also vary in their interpupillary distance, which creates variability in the perceived depth produced by binocular disparity. Furthermore, the capture and viewing geometries also affect the fusibility of the image and possible eyestrain and discomfort (Lambooij, IJsselsteijn, Fortuin, & Heynderickx, 2009). The negative effects of media characteristics on presence are also included in a frequently used sense-of-presence questionnaire (Lessiter, Freeman, Davidoff, & Keogh, 2001) because they have been found to significantly decrease the sense of presence.

In interpreting the results of the present study, it is important to note that in addition to the fundamental difference between the stereoscopic and monoscopic conditions, the negative media characteristics identified above might contribute to the emotions elicited by stereoscopic stimuli. Research has demonstrated that increasing binocular disparity increases the self-reported arousal elicited by stereoscopic images (Kawai et al., 2014), possibly due to negative media characteristics. Large camera bases can produce eye discomfort, which might explain the increased negative valence elicited by the neutral and happy expressions at the high camera base levels found in the present study. Furthermore, crosstalk between images intended for the left and right eyes (Woods, 2012) becomes more visible with larger disparities. For the angry expressions, stereoscopy amplified arousal most at the midrange camera bases, which provided the most natural depth percept. The arousal results for the happy expressions were unexpected and warrant further investigation. Stereoscopy amplified the arousal elicited by happy expressions only at the extreme 15 and 115 mm camera bases, which produced the most unnatural depth percepts. The happy expression result might be explained by self-relevance. Because the happy expressions elicited positive valence, the participants might have felt annoyed by the depth distortion because it reduced the presence elicited by the stimulus and, consequently, reduced self-relevance. Another explanation for the increased arousal elicited by the small camera base is that there might not have been enough three-dimensionality: Users expect stereoscopic images to be fully three-dimensional (Hakala et al., 2014). The negative valence results for the neutral facial expressions are also noteworthy because they indicate that the natural 65 mm camera base elicited significantly less negative valence than the smaller or greater camera bases. This finding is a prime example of the negative effect of unnatural depth on elicited emotional valence. The effect was reversed for the angry expression: The natural 65 mm camera base elicited the highest amplification of negative valence. If the negative valence amplification were due exclusively to the adverse reaction to the negative media

characteristics, the negative valence contrast would have increased monotonically with the camera base as the artefacts and distortions became more noticeable. However, our results show that the negative valence contrast curves resemble U or inverted U-shaped curves (Figure 4), indicating the special significance of the midrange natural camera bases.

In the present study, participants' gender and gender match as well as alexithymia and emotional empathy traits were included as potential confounding factors. Our results indicated that for female participants, angry and neutral expressions elicited stronger negative valence, and happy expressions elicited stronger positive valence. Because both angry and neutral expressions elicited negative emotional reactions (Figure 3), these results can be considered as evidence of higher emotional sensitivity in female participants. This interpretation is confirmed by earlier findings. Compared with males, females have been found to be more accurate in recognizing angry and happy facial expressions (Hall & Matsumoto, 2004), mimic facial expressions more actively (Dimberg & Lundquist, 1990), and report higher positive valence for happy facial expressions and higher negative valence for sad expressions (Wild, Erb, & Bartels, 2001). The gender match results indicate that the expressions of the opposite gender elicit emotions that are more favorable (i.e., less negative). Consistent with these results, previous research has found that pictures of females elicit more positive valence in males than pictures of males (Vrana & Gross, 2004). However, because the stimuli in the present experiment depicted the posed emotional expressions of only two females and two males, gender-specific results must be interpreted with caution. The validity of the present emotional empathy and alexithymia-related results is also limited because these traits were measured in the randomly selected participant sample rather than employed as inclusion criteria. Nevertheless, most of the study results are consistent with previous findings. For example, alexithymia has been associated with difficulties in recognizing both one's own and others' emotions (Bagby et al., 1994), which is consistent with the present finding that angry and neutral (i.e., slightly negative) faces elicited fewer negative experiences and happy faces elicited fewer positive experiences in more alexithymic participants. It should be noted that gender, alexithymia, and empathy did not predict any of the responses elicited by stereoscopic displays.

Further studies are needed to confirm the mechanisms through which stereoscopy amplifies emotions. Including a live condition in the experimental setup in future studies would provide a baseline against which the stereoscopic and monoscopic conditions could be compared. Additionally, the combined effect of stereoscopy and dynamics is the next logical research focus in the endeavor to understand the gap between a picture and the reality that it represents.

Conclusions

The present article presents evidence that stereoscopy amplifies self-reported emotions, but only when depth is natural. Stereoscopy increased the negative valence elicited by angry facial expressions and the positive valence elicited by happy facial expressions. Moreover, for angry facial expressions, the amplification in arousal caused by stereoscopy reached a maximum in the natural depth condition. We propose that stereoscopy amplifies emotional responses through two mechanisms. First, the additional visual cues provided by stereoscopy enhance perception of the facial expression. Second, stereoscopy increases the illusion of nonmediation and thus renders facial expression stimuli more self-relevant to the observer.

Our findings have implications for fields involving pictorial facial expressions, such as human–computer interaction, nonverbal communication, emotion psychology, and

emotion neuroscience. The currently ongoing proliferation of head-mounted virtual reality displays appears to have generated a new boom in stereoscopic displays. Our findings should thus contribute to maximizing the emotional impact of stereoscopic content and increase popular acceptance of the new technology. Furthermore, the study findings provide an opportunity to revisit investigations of emotional facial expressions that were performed using pictorial stimuli.

Acknowledgements

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

References

Ambadar, Z., Schooler, J. W., & Conn, J. F. (2005). Deciphering the enigmatic face the importance of facial dynamics in interpreting subtle facial expressions. *Psychological Science*, *16*, 403–410. doi:10.1111/j.0956-7976.2005.01548.x

Arsalidou, M., Morris, D., & Taylor, M. J. (2011). Converging evidence for the advantage of dynamic facial expressions. *Brain Topography*, *24*, 149–163. doi:10.1007/s10548-011-0171-4

Bagby, R. M., Parker, J. D., & Taylor, G. J. (1994). The twenty-item Toronto Alexithymia Scale—I. Item selection and cross-validation of the factor structure. *Journal of Psychosomatic Research*, *38*, 23–32. doi:10.1016/0022-3999(94)90005-1

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*, 1–48. doi:10.18637/jss.v067.i01

Besel, L. D. S., & Yuille, J. C. (2010). Individual differences in empathy: The role of facial expression recognition. *Personality and Individual Differences*, *49*, 107–112. doi:10.1016/j.paid.2010.03.013

Blascovich, J., Loomis, J., Beall, A. C., Swinth, K. R., Hoyt, C. L., & Bailenson, J. N. (2002). Immersive virtual environment technology as a methodological tool for social psychology. *Psychological Inquiry*, *13*, 103–124. doi:10.1207/S15327965PLI1302_01

Boev, A., Hollosi, D., Gotchev, A., & Egiazarian, K. (2009). Classification and simulation of stereoscopic artifacts in mobile 3DTV content. *Proceedings of SPIE*, *7237*, 72371F–72371F–12. doi:10.1117/12.807185

Bould, E., Morris, N., & Wink, B. (2008). Recognising subtle emotional expressions: The role of facial movements. *Cognition & Emotion*, *22*, 1569–1587. doi:10.1080/02699930801921156

Bradley, M. M., & Lang, P. J. (1994). Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry*, *25*, 49–59.

Burke, D., Taubert, J., & Higman, T. (2007). Are face representations viewpoint dependent? A stereo advantage for generalizing across different views of faces. *Vision Research*, *47*, 2164–2169. doi:10.1016/j.visres.2007.04.018

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1997). Beyond bipolar conceptualizations and measures: The case of attitudes and evaluative space. *Personality and Social Psychology Review: An Official Journal of the Society for Personality and Social Psychology, Inc.*, *1*, 3–25. doi:10.1207/s15327957pspr0101_2

Cacioppo, J. T., Gardner, W. L., & Berntson, G. G. (1999). The affect system has parallel and integrative processing components: Form follows function. *Journal of Personality and Social Psychology*, *76*, 839–855. doi:10.1037/0022-3514.76.5.839

Carr, L., Iacoboni, M., Dubeau, M.-C., Mazziotta, J. C., & Lenzi, G. L. (2003). Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences of the United States of America*, *100*, 5497–5502. doi:10.1073/pnas.0935845100

Devernay, F., & Beardsley, P. (2010). Stereoscopic cinema. In R. Ronfard, & G. Taubin (Eds.), *Image and Geometry Processing for 3-D Cinematography* (Vol. 5, pp. 11–51). Berlin/Heidelberg, Germany: Springerdoi:10.1007/978-3-642-12392-4

Dimberg, U., & Lundquist, L. O. (1990). Gender differences in facial reactions to facial expressions. *Biological Psychology*, *30*, 151–159. doi:10.1016/0301-0511(90)90024-Q

Ekman, P., & Friesen, W. V. (1976). *Pictures of facial affect*. Palo Alto, CA: Consulting Psychologists Press.

Ekman, P., Friesen, W. V., & Hager, J. C. (2002). *Facial action coding system* 2nd ed. Salt Lake City, UT: Research Nexus eBook.

Frith, C. (2009). Role of facial expressions in social interactions. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, *364*, 3453–3458. doi:10.1098/rstb.2009.0142

Goeleven, E., De Raedt, R., Leyman, L., & Verschuere, B. (2008). The karolinska directed emotional faces: A validation study. *Cognition & Emotion*, *22*, 1094–1118. doi:10.1080/02699930701626582

Gronenschild, E. H. B. M., Smeets, F., Vuurman, E. F. P. M., van Boxtel, M. P. J., & Jolles, J. (2009). The use of faces as stimuli in neuroimaging and psychological experiments: A procedure to standardize stimulus features. *Behavior Research Methods*, *41*, 1053–1060. doi:10.3758/BRM.41.4.1053

Hakala, J., Westman, S., Salmimaa, M., Pölönen, M., Järvenpää, T., & Häkkinen, J. (2014). Why 3D cameras are not popular: A qualitative user study on stereoscopic photography acceptance. *3D Research*, *5*, 1–18. doi:10.1007/s13319-013-0004-1

Hall, J. A., & Matsumoto, D. (2004). Gender differences in judgments of multiple emotions from facial expressions. *Emotion*, *4*, 201–206. doi:10.1037/1528-3542.4.2.201

Harwood, N. K., Hall, L. J., & Shinkfield, A. J. (1999). Recognition of facial emotional expressions from moving and static displays by individuals with mental retardation. *American Journal of Mental Retardation: AJMR*, *104*, 270–278.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*, 65–70.

IJsselsteijn, W. A., de Ridder, H., Freeman, J., & Avons, S. E. (2000). Presence: Concept, determinants, and measurement. In *Proceedings of SPIE* (Vol. 31, pp. 520–529). doi:10.1117/12.387188

IJsselsteijn, W. A., de Ridder, H., Hamberg, R., Bouwhuis, D., & Freeman, J. (1998). Perceived depth and the feeling of presence in 3DTV. *Displays*, *18*, 207–214. doi:10.1016/S0141-9382(98)00022-5

Jiang, F., Blanz, V., & O'Toole, A. J. (2009). Three-dimensional information in face representations revealed by identity aftereffects. *Psychological Science*, *20*, 318–325. doi:10.1111/j.1467-9280.2009.02285.x

Kaplan, K. J. (1972). On the ambivalence-indifference problem in attitude theory and measurement: A suggested modification of the semantic differential technique. *Psychological Bulletin*, *77*, 361–372. doi:10.1037/h0032590

Kätsyri, J., & Sams, M. (2008). The effect of dynamics on identifying basic emotions from synthetic and natural faces. *International Journal of Human Computer Studies*, *66*, 233–242. doi:10.1016/j.ijhcs.2007.10.001

Kawai, T., Atsuta, D., Tomiyama, Y., Kim, S., Morikawa, H., Mitsuya, R., & Häkkinen, J. (2014). Disparity modifications and the emotional effects of stereoscopic images. In *Proceedings of SPIE* (Vol. 9011, pp. 901115–1–8). doi:10.1117/12.2042376

Krumhuber, E. G., Kappas, A., & Manstead, A. S. R. (2013). Effects of dynamic aspects of facial expressions: A review. *Emotion Review*, 5, 41–46. doi:10.1177/1754073912451349

Lambooij, M., IJsselsteijn, W., Fortuin, M., & Heynderickx, I. (2009). Visual discomfort and visual fatigue of stereoscopic displays: A review. *Journal of Imaging Science and Technology*, 53, 030201–14. doi:10.2352/J.ImagingSci.Technol.2009.53.3.030201

Lang, P. J. (1980). Behavioral treatment and bio-behavioral assessment: Computer applications. In J. B. Sidowski, J. H. Johnson, & T. A. Williams (Eds.), *Technology in mental health care delivery systems* (pp. 119–137). Norwood, NJ: Ablex Publishing Corporation.

Lessiter, J., Freeman, J., Davidoff, J. B., & Keogh, E. (2001). A cross-media presence questionnaire: The ITC-sense of presence inventory. *Presence*, 10, 282–297. doi:10.1162/105474601300343612

Mehrabian, A. (2000). Manual for the Balanced Emotional Empathy Scale (BEES) (Available from Albert Mehrabian, 1130 Alta Mesa Road, Monterey, CA 93940).

Montagne, B., Kessels, R. P. C., Frigerio, E., De Haan, E. H. F., & Perrett, D. I. (2005). Sex differences in the perception of affective facial expressions: Do men really lack emotional sensitivity? *Cognitive Processing*, 6, 136–141. doi:10.1007/s10339-005-0050-6

N'Diaye, K., Sander, D., & Vuilleumier, P. (2009). Self-relevance processing in the human amygdala: Gaze direction, facial expression, and emotion intensity. *Emotion*, 9, 798–806. doi:10.1037/a0017845

Öhman, A., & Mineka, S. (2001). Fears, phobias, and preparedness: Toward an evolved module of fear and fear learning. *Psychological Review*, 108, 483–522. doi:10.1037/0033-295X.108.3.483

Palermo, R., & Coltheart, M. (2004). Photographs of facial expression: Accuracy, response times, and ratings of intensity. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc.*, 36, 634–638. doi:10.3758/BF03206544

Parker, J. D., Taylor, G. J., & Bagby, R. M. (1993). Alexithymia and the recognition of facial expressions of emotion. *Psychotherapy and Psychosomatics*, 59, 197–202.

Pönkänen, L. M., Alhoniemi, A., Leppänen, J. M., & Hietanen, J. K. (2011). Does it make a difference if I have an eye contact with you or with your picture? An ERP study. *Social Cognitive and Affective Neuroscience*, 6, 486–494. doi:10.1093/scan/nsq068

Pönkänen, L. M., Hietanen, J. K., Peltola, M. J., Kauppinen, P. K., Haapalainen, A., & Leppänen, J. M. (2008). Facing a real person: An event-related potential study. *Neuroreport*, 19, 497–501. doi:10.1097/WNR.0b013e3282f7c4d3

Quené, H., & Van Den Bergh, H. (2004). On multi-level modeling of data from repeated measures designs: A tutorial. *Speech Communication*, 43, 103–121. doi:10.1016/j.specom.2004.02.004

R Core Team. (2015). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Sato, W., & Yoshikawa, S. (2007). Enhanced experience of emotional arousal in response to dynamic facial expressions. *Journal of Nonverbal Behavior*, 31, 119–135. doi:10.1007/s10919-007-0025-7

Takatalo, J., Kawai, T., Kaistinen, J., Nyman, G., & Häkkinen, J. (2011). User experience in 3D stereoscopic games. *Media Psychology*, 14, 387–414. doi:10.1080/15213269.2011.620538

Viinikainen, M., Kätsyri, J., & Sams, M. (2012). Representation of perceived sound valence in the human brain. *Human Brain Mapping*, 33, 2295–2305. doi:10.1002/hbm.21362

Vrana, S. R., & Gross, D. (2004). Reactions to facial expressions: Effects of social context and speech anxiety on responses to neutral, anger, and joy expressions. *Biological Psychology*, 66, 63–78. doi:10.1016/j.biopsycho.2003.07.004

Watson, J. M. (2004). From interpretation to identification: A history of facial images in the sciences of emotion. *History of the Human Sciences*, 17, 29–51. doi:10.1177/0952695104043581

Wild, B., Erb, M., & Bartels, M. (2001). Are emotions contagious? Evoked emotions while viewing emotionally expressive faces: Quality, quantity, time course and gender differences. *Psychiatry Research*, 102, 109–124. doi:10.1016/S0165-1781(01)00225-6

Woods, A. J. (2012). Crosstalk in stereoscopic displays: A review. *Journal of Electronic Imaging*, 21, 040902–1–21. doi:10.1117/1.JEI.21.4.040902

Author Biographies

Jussi Hakala (MSc, Tech.) is finalizing his doctoral thesis on the perceptual and affective effects of stereoscopic imaging technologies at the department of computer science in Aalto University. His interests include imaging technologies, binocular perception, and technology foresight.

Jari Kätsyri (PhD, Cogn Sci and Tech, Helsinki Univ Tech) is a post-doctoral research fellow (Emil Aaltonen Foundation) at Aalto University. His areas of research interest and expertise include emotion studies in both human-human and human–computer interactions.

Jukka Häkkinen received his PhD in experimental psychology from University of Helsinki, Finland. Currently, he is a principal investigator in visual cognition research group at the Institute of Behavioral Sciences, University of Helsinki. He is currently researching various topics in visual cognition.