ORIGINAL PAPER

CrossMark

Auditory Distress Signals Potentiate Attentional Bias to Fearful Faces: Evidence for Multimodal Facilitation of Spatial Attention by Emotion

Joshua M. Carlson[1] · Scott Conger[1] · Jamie Sterr[1]

Abstract
Unimodal emotionally salient visual and auditory stimuli capture attention and have been found to do so cross-modally. However, little is known about the combined influences of auditory and visual threat cues on directing spatial attention. In particular, fearful facial expressions signal the presence of danger and capture attention. Yet, it is unknown whether human auditory distress signals that accompany fearful facial expressions potentiate their capture of attention. It was hypothesized that the capture of attention by fearful faces would be enhanced when co-presented with auditory distress signals. To test this hypothesis, we used a modified multimodal dot-probe task where fearful faces were paired with three sound categories: no sound control, non-distressing human vocalizations, and distressing human vocalizations. Fearful faces captured attention across all three sound conditions. In addition, this effect was potentiated when fearful faces were paired with auditory distress signals. The results provide initial evidence suggesting that emotional attention is facilitated by multisensory integration.

Keywords Multimodal attention · Audiovisual attention · Emotional attention · Nonverbal behavior · Social cognition · Dot-probe

Within our daily lives, a continuous stream of incoming information bombards our sensory systems. The brain is incapable of processing this vast amount of information equally and therefore has developed methods of prioritizing certain environmental stimuli over others. Selectively attending to biologically salient stimuli and/or spatial locations in the environment is one method of prioritization. In particular, stimuli that hold emotional significance automatically capture and hold attention (Torrence et al. 2017). The capture of attention by emotional stimuli has been extensively studied in the visual domain where

✉ Joshua M. Carlson
joshcarl@nmu.edu

[1] Department of Psychology, Northern Michigan University, 1401 Presque Isle Avenue, Marquette, MI 49855, USA

threat signals such as fearful and angry faces, weapons, dangerous animals, and acts of violence (among other threatening/negative stimuli) have all been found to capture attention compared to emotionally neutral stimuli (Carlson et al. 2009; Cooper and Langton 2006; Fox 2002; Koster et al. 2004; Macleod et al. 1986; Salemink et al. 2007). In addition, appetitive visual signals such as happy faces, pictures of babies, and other affiliative stimuli also capture attention (Brosch et al. 2008b, 2007; Elam et al. 2010; Kreta et al. 2016; Torrence et al. 2017). Typically, emotional influences on selective spatial attention are studied within-modality in a unimodal manner (i.e., visual emotional cues enhance visual processing); however, emotion can also facilitate attention across sensory modalities. Specifically, auditory threat signals enhance visual processing (Brosch et al. 2008a, 2009; Zeelenberg and Bocanegra 2010) and visual threat signals enhance auditory processing (Harrison and Woodhouse 2016; Selinger et al. 2013).

Yet, the extent to which *multimodal* emotional signals enhance attention relative to unimodal emotional signals is unknown. Within our natural environment, auditory and visual inputs are seamlessly integrated to form a unified audiovisual percept. Therefore, it is reasonable to expect that converging audiovisual signals of emotion would have a greater effect on guiding attention relative to unimodal signals of emotion. For example, fearful facial expressions are important social signals of potential environmental threat that capture attention. In the natural environment, fearful facial expressions are often paired with auditory distress signals (e.g., screams). One possibility is that such multimodal threat signals more effectively capture attention relative to unimodal signals. An alternative possibility, is that any threat (or emotional) signal elicits an "all-or-none" response and therefore multimodal and unimodal signals capture attention equally. To test between these two possibilities, we used a multimodal audiovisual dot-probe task to explore the degree to which auditory distress signals (e.g., screams) modulate the attention grabbing effects of fearful faces. We hypothesized that audiovisual threat signals (i.e., fearful face + scream) would produce a greater capture of attention relative to unimodal threat signals (i.e., fearful face only).

Method

Participants

Sixty-one undergraduate students (female = 34, right handed = 56) between the ages of 18 and 33 ($M = 20.20$, $SD = 2.20$) participated in the study.[1] After review of box and whisker plots for each condition, three individuals were identified as outliers for having consistently slower reaction times across conditions and were therefore excluded from analyses.[2] The remaining sample contained 58 participants between 18 and 33 years old ($M = 20.24$, $SD = 2.25$, female = 31, right handed = 53). The Northern Michigan University Institutional Review Board approved all aspects of the study. Participants received course credit for their participation.

---

[1] A sample size estimate was based on the smallest effect size (i.e., $\eta_p^2 = 0.08$) reported in an earlier study of crossmodal affective (auditory) attention (Brosch et al., 2008). Using G*Power 3.1.9.2 with $\eta_p^2 = 0.08$, $\alpha = 0.05$, and power = 0.95 it was determined that an $N \geq 52$ would be needed to detect similarly sized effects.

[2] Note that all effects reported below were also significant in the full sample of $N = 61$.
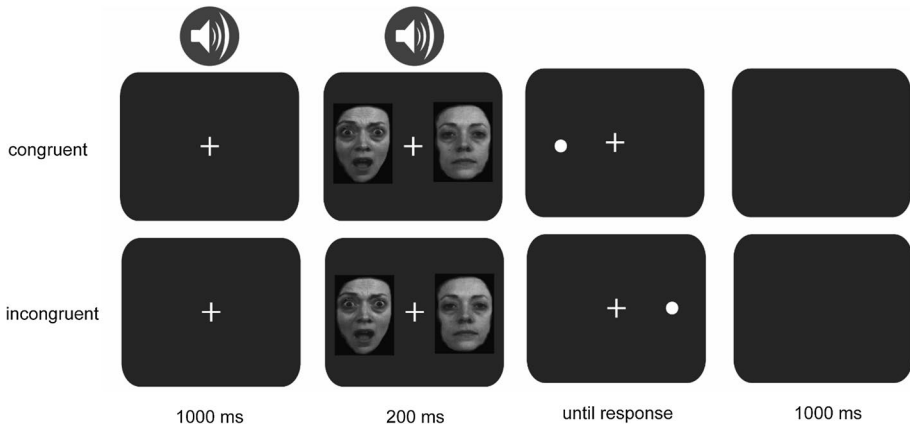
Fig. 1 Examples of congruent and incongruent trials in the dot-probe task. The task contained three sound conditions (no sound, non-distress sound, and distress sound), which occurred during the fixation and face cues for 1200 ms

Stimuli and Testing Apparatus

Four fearful and neutral grayscale facial identities were used as the visual stimuli (from Gur et al. 2002). Ten nonverbal auditory files were selected from the International Affective Digitized Sounds (i.e., IADS) database (Bradley and Lang 2007), which included human distress sounds (e.g., screams) as well as human non-distress sounds (e.g., laughing, yawning, and whistling).[3] Auditory files from the IADS were originally 6000 ms in duration. Using Audacity sound editing software (http://audacity.sourceforge.net/), these files were trimmed to 1200 ms segments[3]. The resulting stimuli used in the experiment contained nine unique human distress sounds (M_{dB}=74.33, M_{Hz}=1135.25) and nine unique human non-distress sounds (M_{dB}=73.11, M_{Hz}=573.02) equated on decibel level (p>.2), but different in their fundamental frequency (p=.02). Given that distress and non-distress sounds differed in fundamental frequency, we explored the relationship between frequency and attentional bias. No consistent relationship was observed between these variables (r=−.2, p>.4; see Table 2 in the Appendix for details and additional comparisons). In a separate sample (N=10), the trimmed IADS auditory files for human distress sounds ($M_{valence}$=2.62, $M_{arousal}$=6.83) were rated as more unpleasant/negative (p=.0001) and more exciting/arousing (p=.001) than the human non-distress sounds ($M_{valence}$=5.69, $M_{arousal}$=4.51). The task was programmed using E-Prime2 (Psychology Software Tools, Pittsburg, PA). Visual stimuli were displayed on a 60 Hz 16″ LCD computer monitor and auditory stimuli were presented through Sony MDR-ZX110NC noise canceling headphones.

Dot-Probe Task

As displayed in Fig. 1, each trial started with a white fixation cue (+) centered on a black background for 1000 ms. Two faces were then presented simultaneously, one on each side of the fixation cue, for 200 ms. We chose a relatively short stimulus duration for our visual

---

[3] IADS files 275, 276, 277, 279, and 285 were used to create our distress sounds resulting in 275.1, 276.1, 276.2, 277.1, 277.2, 279.1, 279.2, 285.1, and 285.2. Files 221, 226, 230, 262, and 270 were used to create our non-distress sounds resulting in 221.1, 226.1, 226.2, 230.1, 230.2, 262.1, 262.2, 262.3, 270.1.

cues based on previous research indicating that shorter (i.e., < 300 ms) presentation times and stimulus onset asynchronies produce stronger attention effects for fearful faces in the dot-probe task (Torrence et al. 2017). Faces subtended 5° × 7° of the visual angle and were separated by 14° of the visual angle. After face offset, a target dot appeared on the left or right side of the screen and remained until a response was made with an E-Prime Chronos response box. Participants indicated left-sided targets by pressing the first, leftmost button using their right index finger and indicated right-sided targets by pressing the second button using their right middle finger. Participants were instructed to focus on the central fixation cue throughout the experiment and respond to the targets as quickly as possible.

On one third of the trials, no auditory stimuli were presented. Distress and non-distress sounds were equally presented on the other two-thirds of trials (i.e., a third each). Auditory stimuli were presented binaurally[4] for 1200 ms starting at the onset of the fixation cue and terminating on the offset of the face pair. Similar to previous research (Brosch et al. 2008a), we used a longer duration time for the auditory stimuli as the emotional significance communicated by the nonverbal auditory expression takes time to unfold.

The task included an equal number of congruent (dot on the same side as the fearful face) and incongruent (dot on the same side as the neutral face) trials in which the fearful face was displayed with equal probability to the left or right visual field. Faster responses on congruent compared to incongruent trials is representative of an attentional bias to fearful faces. The task consisted of 12 possible trial combinations depending on a combination of sound type, congruency, and the visual field in which the fearful face occurred (i.e., 2 × 2 × 3). There were 480 total trials: 160 no sound trials (80 congruent and 80 incongruent), 160 non-distress sound trials (80 congruent and 80 incongruent), and 160 distress sound trials (80 congruent and 80 incongruent). The four fearful and neutral face stimuli were each presented 120 times (approximately 20 times each for congruent and incongruent trials across the three sound conditions). Each of the nine distress and non-distress sounds were presented approximately nine times in both congruent and incongruent trials. The task contained four blocks in which each of the 12 trial combinations were randomly presented 10 times for 120 trials per block. After each block, participants received feedback about their average accuracy and mean response time to encourage accurate and quick responses. All trials were presented in a unique randomized order across participants.

Data Analysis

Data were filtered to only include correct responses between 150 and 750 ms post-target onset to exclude premature and delayed responses (Carlson and Mujica-Parodi 2015; Carlson and Reinke 2014; Carlson et al. 2016; Torrence et al. 2017). After filtering, 96.60% of the data were available for use in analyses. To test the effects of the within subjects variables congruency (congruent vs. incongruent) and sound type[5] (no sound vs. non-distress sound vs. distress sound), a 2 × 3 repeated measures analysis of variance (ANOVA) was conducted with reaction time as the dependent variable. When appropriate, significant

---

[4] Note that we used a binaural presentation of auditory stimuli to ensure that the effects of the auditory stimulus were only modulatory, as they themselves did not contain any spatial information. That is, any potential effects of the auditory stimulus on enhancing fearful face-elicited attention could not be due to the direct influence of the auditory stimulus on spatial attention (Brosch et al. 2008a, 2009).

[5] Note that the variable sound type represents a continuum of saliency from no sound to non-distress sound to distress sound. Therefore, the levels of sound type were organized along this continuum in our analyses.

main and interaction effects were explored with single-tailed Bonferroni corrected pairwise comparisons.

Results

There was a main effect of congruency, $F(1, 57) = 93.32$, $p < .001$ $\eta_p^2 = .63$, where reaction times were faster on congruent compared to incongruent trials (see Table 1 for Means, SDs, and 95% Confidence Intervals). There was also a main effect of sound type, $F(1, 114) = 81.10$, $p < .001$, $\eta_p^2 = .59$. Follow up pairwise comparisons indicate that distress sound trials were significantly faster than non-distress sound ($p = .01$) and no sound trials ($p < .001$). Non-distress sound trials were also faster than no sound trials ($p < .001$). As this pattern would suggest, there was a strong linear trend where reaction times decreased from the least salient sound condition (i.e., no sound) to the most salient sound condition (i.e., distress sound), F_{Linear} $(1, 57) = 116.78$, $p < .001$, $\eta_p^2 = .67$. The congruency by sound type interaction did not reach traditional levels of statistical significance, $F(2, 114) = 2.59$, $p = .08$, $\eta_p^2 = .04$. However, a significant linear trend for the congruency by sound type interaction was observed, F_{Linear} $(1, 57) = 4.79$, $p = .03$, $\eta_p^2 = .08$. As can be seen in Fig. 1 and Table 1, congruent trials were faster than incongruent trials for the no sound condition ($p < .001$), the non-distress sound condition ($p < .001$), and the distress sound condition ($p < .001$). This effect linearly increased from the least to the most salient sound conditions (i.e., no sound to distress sound). In particular, congruent trials were significantly faster in the distress sound condition compared to both the non-distress sound ($p = .005$) and no sound conditions ($p < .001$). Congruent trials in the non-distress sound condition were faster than the no sound condition ($p < .001$). On the other hand, incongruent trials were significantly slower in the no sound condition compared to the non-distress ($p < .001$) and distress sound conditions ($p < .001$); but, non-distress and distress sound types did not differ on incongruent trials ($p = .82$).

Discussion

Non-distress and distress sounds produced a non-specific facilitation of reaction time, which was greatest for distress sounds. More importantly, as hypothesized, we found that although fearful faces captured attention across all conditions, this effect was largest in the presence of distress sounds. Specifically, quicker responses on congruent trials drove the effect—suggesting that auditory distress signals facilitate the initial capture of attention.

Our results are the first to indicate that multimodal audiovisual signals of threat (i.e., fearful faces + distress sounds) produce a relatively greater enhancement of attention compared to unimodal threat signals (i.e., fearful faces alone). These findings add to previous research indicating that fearful faces and other unimodal emotional stimuli capture attention (Brosch et al. 2008b, 2007; Carlson et al. 2009; Carlson and Mujica-Parodi 2015; Carlson and Reinke 2008; Cooper and Langton 2006; Elam et al. 2010; Fox 2002; Koster et al. 2004; Macleod et al. 1986; Salemink et al. 2007; Torrence et al. 2017). The findings also add to work demonstrating that emotional attention operates in a cross-modal nature such that emotional signals in one sensory modality enhance attention and perception in other sensory modalities (Brosch et al. 2008a, 2009; Harrison and Woodhouse 2016; Selinger et al. 2013; Zeelenberg and Bocanegra 2010). Yet, until now, it was unknown

Table 1 Means, SDs, and 95% CIs for congruency by sound type in ms

	No sound	Non-distress sound	Distress sound	Marginal means
Congruent	340.46, 31.04 (332.30–348.62)	329.12, 32.42 (320.59–337.64)	325.57, 30.54 (317.54–333.60)	332.72, 30.81 (323.61–339.83)
Incongruent	349.83, 28.27 (342.40–357.26)	339.97, 30.15 (332.04–347.90)	338.57, 28.46 (331.09–346.06)	342.79, 28.23 (335.36–350.23)
Marginal means	345.15, 29.14 (337.47–352.82)	334.54, 30.74 (326.45–342.64)	332.07, 28.91 (324.46–339.69)	337.25, 29.21 (329.56–344.95)

as to whether multiple sources of converging sensory signals of emotion would further enhance attention. We found a nonspecific enhancement in reaction time by (binaural) non-verbal human sounds, which was greatest for distress signals—suggesting a general alert-ing effect. Moreover, we found that human distress signals facilitate the spatially specific capture of attention by fearful faces. This effect was attributed to faster responses on con-gruent trials for distress sounds compared to all other conditions, which suggests that the initial capture of attention by fearful faces was enhanced by the presence of an auditory distress signal. Thus, multimodal threat signals more effectively capture attention relative to unimodal signals. Past research has shown that unimodal emotional cues not only facili-tate reaction time, but also modulate electrocortical processing in occipitotemporal areas (Brosch et al. 2009; Carlson and Reinke 2010; Torrence and Troup 2017). An interesting avenue for future studies measuring attention to multimodal human distress signals would be to assess the degree to which auditory distress cues paired with fearful faces enhance the amplitude of the face-selective N170 event-related potential above and beyond that of fearful faces alone (Carlson and Reinke 2010; Torrence and Troup 2017).

Our results add to an established line of research indicating that audiovisual face and voice stimuli lead to better recognition of emotion than unimodal stimuli (for review see Gerdes et al. 2014). The facilitation of emotion recognition by multimodal face and voice stimuli is independent of endogenous attention (Collignon et al. 2008; Focker et al. 2011; Vroomen et al. 2001). Our results are consistent with the notion that multimodal face and voice stimuli automatically enhance emotional processing, but further indicate that this effect is not only true for facilitated emotion recognition, but also exogenous selective vis-ual attention. Beyond emotion recognition and attention, it is unclear what other aspects of information processing are enhanced by multimodal cues of emotion. Given the ubiquity of multimodal audiovisual sensory inputs and their importance in social cognition (de Gelder and Bertelson 2003), it would be expected that additional aspects of information process-ing (at least those relevant to social cognition) are enhanced by multimodal stimuli relative to unimodal stimuli.

Note that by using only fearful and neutral faces with distress and non-distress sounds the generalizability of our results to other emotional stimuli is limited.[6] In par-ticular, it is unclear if such effects would also exist for happy facial expressions and joyful sounds or other types of audiovisual emotional stimuli. Furthermore, it is unclear whether auditory and visual information originating from a single source (e.g., a fearful face + scream) has an advantage over multimodal inputs arising from multiple sources (e.g., a fearful face + growling animal). Future research should focus on exploring these unknown aspects of multimodal emotional attention. Additional limitations of our study include the restricted number of exemplars of human distress and non-distress sounds in the IADS database and the difference in frequency between these sounds (see sup-plementary material and analyses for potential frequency-related effects). Although sup-plemental analyses indicate there is no relationship between fundamental frequency and attentional bias, future research would benefit from using a larger database of nonverbal human sounds matched on frequency. The restricted number of non-distressing human sounds in the IADS resulted in the grouping of laughing, yawning, and whistling, which differ in their emotional valence and arousal. Given this, the potentiation of attentional bias to fearful faces by auditory distress signals could theoretically be explained by a

---

[6] Note that this limitation was due to our prioritization of establishing a multimodal effect using a visual stimulus that has consistently been shown to capture attention in our lab.

general category congruency effect. However, if this were the case, we would expect to see greater attentional bias *toward neutral faces* during the neutral/boring sounds (i.e., yawns). Yet, all non-distress sounds were accompanied by an attentional bias to the fearful faces of approximately similar magnitude—suggesting that the potentiation of attentional bias to fearful faces by human distress sounds is not driven by a category congruency effect (see Table 2 in the Appendix). Finally, compared to the large effect size observed for the attentional capture by fearful faces, the potentiation of this effect by distress sounds was modest. Thus, although multimodal emotional signals enhance attention, unimodal emotional stimuli are themselves strong signals of environmental salience that effectively capture observers' attention.

In conclusion, fearful faces captured attention across all sound conditions. Importantly, however, the attentional capture by fearful faces was largest in the presence of distress sounds—providing evidence for a multimodal facilitation of spatial attention by emotion.

Appendix

See Table 2.

Table 2 Mean characteristics of threat and non-threat sounds used in dot-probe task

Sound description	File number	Type	DB	Hz	Mean attention bias (95% CI)	Valence	Arousal
Laughter	221.1	Non-threat	73	1225	10.81 (2.58, 19.04)	5.9	5.10
Laughter	226.1	Non-threat	70	232.11	16.54 (8.52, 24.55)	6.2	5.40
Laughter	226.2	Non-threat	69	259.41	8.92 (1.29, 16.54)	6.2	5.70
Laughter	230.1	Non-threat	74	525	11.11 (2.93, 19.30)	5.8	5.30
Laughter	230.2	Non-threat	70	474.19	9.26 (0.59, 17.92)	6.5	5.70
Yawn	262.1	Non-threat	77	341.86	8.03 (−0.60, 16.67)	5	2.80
Yawn	262.2	Non-threat	77	376.92	4.24 (−5.06, 13.53)	5.2	3.30
Yawn	262.3	Non-threat	74	344.53	17.91 (9.09, 26.73)	5	3.00
Whistle	270.1	Non-threat	74	1378.13	11.01 (2.80, 19.21)	5.2	4.30
Screaming	275	Threat	73	832.08	0.57 (−9.24, 10.38)	2.3	7.40
Screaming	276.1	Threat	74	1837.5	19.47 (10.92, 28.03)	2.8	6.90
Screaming	276.2	Threat	72	1191.89	13.34 (4.02, 22.66)	2.6	6.60
Screaming	277.1	Threat	74	1917.39	13.78 (5.72, 21.83)	2.9	6.90
Screaming	277.2	Threat	76	1102.5	11.27 (2.51, 20.02)	2.8	7.10
Screaming/yelling	279.1	Threat	74	801.82	16.96 (9.91, 24.00)	2.8	6.60
Screaming/yelling	279.2	Threat	74	735	20.65 (10.40, 30.89)	2.3	6.40
Screaming/yelling	285.1	Threat	75	329.1	10.51 (2.46, 18.57)	2.4	6.80
Screaming	285.2	Threat	77	1470	9.21 (1.02, 17.40)	2.7	7.10

References

Bradley, M. M., & Lang, P. J. (2007). *The international affective digitized sounds (2nd ed.; IADS-2): Affective ratings of sounds and instruction manual*. Gainesville, FL: University of Florida.

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2008a). Behold the voice of wrath: Cross-modal modulation of visual attention by anger prosody. *Cognition, 106*(3), 1497–1503. https://doi.org/10.1016/j.cognition.2007.05.011.

Brosch, T., Grandjean, D., Sander, D., & Scherer, K. R. (2009). Cross-modal emotional attention: Emotional voices modulate early stages of visual processing. *Journal of Cognitive Neuroscience, 21*(9), 1670–1679. https://doi.org/10.1162/jocn.2009.21110.

Brosch, T., Sander, D., Pourtois, G., & Scherer, K. R. (2008b). Beyond fear: Rapid spatial orienting toward positive emotional stimuli. *Psychological Science, 19*(4), 362–370. https://doi.org/10.1111/j.1467-9280.2008.02094.x.

Brosch, T., Sander, D., & Scherer, K. R. (2007). That baby caught my eye… attention capture by infant faces. *Emotion, 7*(3), 685–689. https://doi.org/10.1037/1528-3542.7.3.685.

Carlson, J. M., Fee, A. L., & Reinke, K. S. (2009). Backward masked snakes and guns modulate spatial attention. *Evolutionary Psychology, 7*(4), 527–537.

Carlson, J. M., & Mujica-Parodi, L. R. (2015). Facilitated attentional orienting and delayed disengagement to conscious and conconscious fearful faces. *Journal of Nonverbal Behavior, 39*(1), 69–77. https://doi.org/10.1007/s10919-014-0185-1.

Carlson, J. M., & Reinke, K. S. (2008). Masked fearful faces modulate the orienting of covert spatial attention. *Emotion, 8*(4), 522–529. https://doi.org/10.1037/a00126532008-09984-008.

Carlson, J. M., & Reinke, K. S. (2010). Spatial attention-related modulation of the N170 by backward masked fearful faces. *Brain and Cognition, 73*(1), 20–27. https://doi.org/10.1016/j.bandc.2010.01.007S0278-2626(10)00017-5.

Carlson, J. M., & Reinke, K. S. (2014). Attending to the fear in your eyes: Facilitated orienting and delayed disengagement. *Cognition and Emotion, 28*(8), 1398–1406. https://doi.org/10.1080/02699931.2014.885410.

Carlson, J. M., Torrence, R. D., & Vander Hyde, M. R. (2016). Beware the eyes behind the mask: The capture and hold of selective attention by backward masked fearful eyes. *Motivation and Emotion, 40*(3), 498–505. https://doi.org/10.1007/s11031-016-9542-1.

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., et al. (2008). Audio-visual integration of emotion expression. *Brain Research, 1242,* 126–135. https://doi.org/10.1016/j.brainres.2008.04.023.

Cooper, R. M., & Langton, S. R. (2006). Attentional bias to angry faces using the dot-probe task? It depends when you look for it. *Behaviour Research and Therapy, 44*(9), 1321–1329.

de Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Science, 7*(10), 460–467.

Elam, K. K., Carlson, J. M., DiLalla, L. F., & Reinke, K. S. (2010). Emotional faces capture spatial attention in 5-year-old children. *Evolutionary Psychology, 8*(4), 754–767.

Focker, J., Gondan, M., & Roder, B. (2011). Preattentive processing of audio-visual emotional signals. *Acta Psychologica (Amst), 137*(1), 36–47. https://doi.org/10.1016/j.actpsy.2011.02.004.

Fox, E. (2002). Processing emotional facial expressions: The role of anxiety and awareness. *Cognitive, Affective, and Behavioral Neuroscience, 2*(1), 52–63.

Gerdes, A. B., Wieser, M. J., & Alpers, G. W. (2014). Emotional pictures and sounds: A review of multimodal interactions of emotion cues in multiple domains. *Frontiers in Psychology, 5,* 1351. https://doi.org/10.3389/fpsyg.2014.01351.

Gur, R. C., Sara, R., Hagendoorn, M., Marom, O., Hughett, P., Macy, L., et al. (2002). A method for obtaining 3-dimensional facial expressions and its standardization for use in neurocognitive studies. *Journal of Neuroscience Methods, 115*(2), 137–143.

Harrison, N. R., & Woodhouse, R. (2016). Modulation of auditory spatial attention by visual emotional cues: Differential effects of attentional engagement and disengagement for pleasant and unpleasant cues. *Cognitive Processing, 17*(2), 205–211. https://doi.org/10.1007/s10339-016-0749-6.

Koster, E. H., Crombez, G., Verschuere, B., & De Houwer, J. (2004). Selective attention to threat in the dot probe paradigm: Differentiating vigilance and difficulty to disengage. *Behaviour Research and Therapy, 42*(10), 1183–1192.

Kreta, M. E., Jaasmab, L., Biondac, T., & Wijnend, J. G. (2016). Bonobos (Pan paniscus) show an attentional bias toward conspecifics' emotions. *Proceedings of the National Academy of Sciences of the United States of America*. https://doi.org/10.1073/pnas.1522060113.

Macleod, C., Mathews, A., & Tata, P. (1986). Attentional bias in emotional disorders. *Journal of Abnormal Psychology, 95*(1), 15–20.

Salemink, E., van den Hout, M. A., & Kindt, M. (2007). Selective attention and threat: Quick orienting versus slow disengagement and two versions of the dot probe task. *Behaviour Research and Therapy, 45*(3), 607–615.

Selinger, L., Dominguez-Borras, J., & Escera, C. (2013). Phasic boosting of auditory perception by visual emotion. *Biological Psychology, 94*(3), 471–478. https://doi.org/10.1016/j.biopsycho.2013.09.004.

Torrence, R. D., & Troup, L. J. (2017). Event-related potentials of attentional bias toward faces in the dot-probe task: A systematic review. *Psychophysiology*. https://doi.org/10.1111/psyp.13051.

Torrence, R. D., Wylie, E., & Carlson, J. M. (2017). The time-course for the capture and hold of visuospatial attention by fearful and happy faces. *Journal of Nonverbal Behavior*. https://doi.org/10.1007/s10919-016-0247-7.

Vroomen, J., Driver, J., & de Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, Affective, and Behavioral Neuroscience, 1*(4), 382–387.

Zeelenberg, R., & Bocanegra, B. R. (2010). Auditory emotional cues enhance visual perception. *Cognition, 115*(1), 202–206. https://doi.org/10.1016/j.cognition.2009.12.004.