When emotional valence modulates audiovisual integration

**Mario Maiworm · Marina Bellantoni · Charles Spence ·
Brigitte Röder**

Abstract We constantly integrate the information that is
available to our various senses. The extent to which the
mechanisms of multisensory integration are subject to the
influences of attention, emotion, and/or motivation is cur-
rently unknown. The "ventriloquist effect" is widely as-
sumed to be an automatic crossmodal phenomenon,
shifting the perceived location of an auditory stimulus to-
ward a concurrently presented visual stimulus. In the present
study, we examined whether audiovisual binding, as indi-
cated by the magnitude of the ventriloquist effect, is influ-
enced by threatening auditory stimuli presented prior to the
ventriloquist experiment. Syllables spoken in a fearful voice
were presented from one of eight loudspeakers, while sylla-
bles spoken in a neutral voice were presented from the other
seven locations. Subsequently, participants had to localize
pure tones while trying to ignore concurrent visual stimuli
(both the auditory and the visual stimuli here were emotion-
ally neutral). A reliable ventriloquist effect was observed.
The emotional stimulus manipulation resulted in a reduction
of the magnitude of the subsequently measured ventriloquist
effect in both hemifields, as compared to a control group
exposed to a similar attention-capturing, but nonemotional,
manipulation. These results suggest that the emotional sys-
tem is capable of influencing multisensory binding process-
es that have heretofore been considered automatic.

Keywords Multisensory processing · Spatial localization

M. Maiworm (✉) · M. Bellantoni · B. Röder
Department of Psychology, University of Hamburg,
Hamburg, Germany
e-mail: mario.maiworm@uni-hamburg.de

C. Spence
Department of Experimental Psychology, University of Oxford,
Oxford, England

We live in a multisensory world, meaning that for most
perceptual problems, information from multiple modalities
is available at the same time. The question of *how* to
combine multisensory inputs has given rise to a number of
theories over the years (e.g., Ernst & Bülthoff, 2004; Welch,
1999; Welch & Warren, 1980). Currently, it is commonly
assumed that a relation exists between how appropriate a
sensory modality is for a given perceptual task (or, more
precisely, how reliable the perceptual input from that senso-
ry channel is) and the degree of influence that this modality
has on the ensuing multisensory perceptual estimate. For
example, for the task of localizing an audiovisual target, we
would expect vision to dominate the multisensory estimate,
because the visual system tends to be more reliable than the
auditory system for spatial tasks (e.g., Alais & Burr, 2004).

This form of visual dominance sometimes results in
illusory percepts such as the "ventriloquist effect" (see
Bertelson & de Gelder, 2004, for a review). If participants
have to localize tones while trying to ignore concurrently
presented, spatially displaced visual stimuli, their responses
are usually biased toward the to-be-ignored visual stimuli.
Abundant evidence has suggested that ventriloquism is an
automatic, genuinely perceptual phenomenon (e.g., Bertelson
& Aschersleben, 1998; Bertelson, Pavani, Ladavas, Vroomen,
& de Gelder, 2000; Bertelson, Vroomen, de Gelder, & Driver,
2000; Bonath et al., 2007; Driver, 1996; Vroomen, Bertelson,
& de Gelder, 2001). This assumption is supported by the
observation that the audiovisual ventriloquist effect is so
robust that it still occurs even if the participants are (or
become) aware of the spatial discrepancy between the audito-
ry and visual stimuli (see Radeau & Bertelson, 1974). How-
ever, more recently, some authors have questioned the full
automaticity of this crossmodal binding effect and have sug-
gested that attention might at least have a modulating influ-
ence (Fairhall & Macaluso, 2009; Röder & Büchel, 2009).

Here, we tested whether or not crossmodal binding is an automatic process. In order to maximize the likelihood of interfering with a presumably automatic crossmodal binding process, as assessed with the ventriloquist illusion, we used emotional stimuli.

The emotional significance of stimuli has often been found to enhance attentional, perceptual, and cognitive processes (Vuilleumier, 2005). For instance, the attentional blink tends to be reduced for aversive emotional stimuli: That is, the second target in a rapidly presented series of stimuli is masked to a lesser extent by a preceding target when the second target is emotionally relevant (as compared to emotionally neutral; Anderson, 2005). This finding exemplifies the fact that the emotional valence of a stimulus can influence processes of attentional selection. Similarly, it has been reported that visual search times are lower for emotional than for neutral stimuli (Fox, 2002), and that the presence of emotional stimuli can interfere with the processing of neutral stimuli. For example, response times to neutral stimuli are prolonged in the context of competing aversive stimuli, as compared to a context in which the competing stimuli are neutral (Fox, Russo, & Dutton, 2002). Finally, when participants are presented with words that are emotionally meaningful, their response times to name the color in which the word is written are increased, as compared to when they respond to emotionally neutral words (see Williams, Mathews, & MacLeod, 1996). This phenomenon, known as the "emotional Stroop effect," demonstrates that the emotional content of words is automatically processed and can interfere with a primary task such as color naming. Given the many findings regarding the effects of emotional significance on the processing of unimodal stimuli, our goal in the present study was to manipulate the emotional valence of stimuli in a ventriloquist situation in order to test whether the emotional valence of stimuli is capable of influencing crossmodal binding, which would, in turn, argue against the ventriloquist illusion being fully automatic.

There are several different possibilities as to how emotional stimuli might influence audiovisual binding. For instance, it has previously been shown that aversive conditioning can result in the sharpening of the frequency tuning of auditory neurons (Bakin & Weinberger, 1990; Weinberger, 2004). Likewise, it might be possible that spatial tuning functions are shifted or sharpened, allowing for a more precise coding of auditory space. Following the idea that the more reliable input influences the multisensory percept more (Ernst & Bülthoff, 2004; Welch, 1999; Welch & Warren, 1980), it would be predicted that such an increased reliability of the auditory channel should result in a higher weighting for auditory inputs as compared to visual inputs in the ventriloquist situation (see, e.g., Alais & Burr, 2004). As a consequence, a reduced magnitude of the ventriloquist effect would be predicted. This means that changes of crossmodal binding—for instance, in the size of the ventriloquist effect—could be accounted for by changes in unimodal, rather than in genuinely crossmodal, processes.

Alternatively, the process of crossmodal binding per se might be what is affected by an emotional manipulation of one stimulus modality. In other words, the emotional manipulation might not, or might to a lesser degree, affect unimodal processes, but rather might change the stimulus processing at a level at which audiovisual integration takes place. Indeed, crossmodal binding can be influenced by top-down manipulations such as variations in attentional load (see Talsma, Senkowski, Soto-Faraco, & Woldorff, 2010, for a review) and stimulus context (Kitagawa & Spence, 2005).

To investigate a possible influence of emotion on crossmodal binding, in the main experiment we utilized a ventriloquist setting with simple tones and lights (see, e.g., Bertelson & Aschersleben, 1998; Vroomen & de Gelder, 2004). To avoid trivial stimulus-driven effects, we used a multistep paradigm: The main experiment was preceded by an emotional-learning phase in which bisyllabic pseudowords, spoken in a fearful voice, were presented from one of four loudspeakers in one hemifield, while emotionally neutral pseudowords were presented from the remaining loudspeakers of the same hemifield and from all four loudspeakers in the opposite hemifield. In the main experiment, emotionally neutral tones were presented together with visual stimuli (created using laser beams) in either the same or a different location with respect to the tone. The participants' task was to ignore the lights and to try and localize the sounds. Thus, we created a typical ventriloquist situation, and we expected the auditory stimuli to be localized toward the concurrently presented visual stimuli. The main aim of the present study was to test whether the size of the ventriloquist effect would be influenced by any emotional conditioning that had taken place during the emotional-learning phase of the experiment. As outlined above, different mechanisms could cause such a modulation of crossmodal binding. First, if emotional learning results in a sharpening of auditory spatial representations, we would expect that the auditory stimulus—now being relatively more reliable—would receive a higher weighting when integrated with the (unchanged) visual stimulus, resulting in a reduced magnitude of the ventriloquist effect (Alais & Burr, 2004; Ernst & Bülthoff, 2004). On the other hand, it might be hypothesized that the crossmodal binding process would be modulated by the emotional manipulation, while unimodal performance would remain unchanged. In order to disentangle these two possible explanations, we introduced unimodal preexperiments and additional conditions (unimodal auditory and congruent audiovisual stimuli) in the main experiment. Moreover, a control group with an attention-demanding learning phase but without the presentation of any emotional stimuli was included. Thus, any group

differences observed during the main experiment could unequivocally be attributed to the emotional manipulation in the preceding learning phase.

Finally, we tested the spatial selectivity of any effect induced by emotional learning. If an altered ventriloquist effect were to be seen only at the emotionally conditioned loudspeaker position, it could be concluded that the emotional significance of the stimuli in the emotional-learning phase of the study had been associated with the location from which the aversive stimuli had been presented. By contrast, if the effects of the emotional manipulation were to be seen at all speaker locations, this would suggest that intermodal selection was affected: The presentation of aversive stimuli in the auditory modality might result in a higher separation between the modality channels, and thus in a reduced ventriloquist effect, irrespective of the spatial location of the stimuli.

Method

Participants

A group of 36 healthy adults (28 female, 8 male), ranging in age from 19 to 28 years ($M = 22.7$ years), took part in this study. The majority were students at the University of Hamburg, and they received either course credit or monetary compensation in return for their participation in the study. All of the participants had normal hearing and normal or corrected-to-normal vision, and all were naïve as to the purpose of the study. The participants gave informed consent prior to their participation, and the whole experimental procedure took approximately 2 h to complete.

Materials and stimuli

The experiment took place in a darkened, anechoic chamber. The participants were seated in front of a device capable of delivering audiovisual stimuli and allowing them to point toward the apparent source of the stimuli by means of a pointer with two degrees of freedom (azimuth and elevation). Eight loudspeakers (two-way speakers, 80-mm cone, 11-mm tweeter dome, model type ConceptC Satellit, Teufel GmbH, Berlin, Germany) were attached to the device in such a way that they extended on a semicircle in front of the participant at ear level. The loudspeakers were positioned on the horizontal plane, extending 7.2 °, 14.4 °, 21.6 °, and 28.8 ° to the left and right of the midline. A chinrest was used to immobilize each participant's head. The distance between the head and any one of the loudspeakers was ~90 cm. The loudspeakers were hidden from the participants' view by an acoustically transparent black curtain that extended in a semicircle in front of the loudspeaker array. Brief tones (1000-Hz sine waves with a

duration of 10 ms) or meaningless bisyllabic pseudowords ("dedu," "none," and "raro," pronounced with a German accent) were played from one of the loudspeakers at ~64 dB (A) in the test and emotional-learning phases, respectively. The pseudowords were spoken by professional male and female actors in either a fearful or a neutral voice. These particular experimental stimuli had previously been developed and validated (see Klinge, Röder, & Büchel, 2010).

The visual stimuli were provided by two laser pointers attached to a rack. The red laser beams were projected onto the black curtain in front of the loudspeakers. One laser pointer was used to provide a static fixation point at the midline (0 °). The other one was angled by means of a stepper motor, allowing for the presentation of visual stimuli at any location along the plane on which the loudspeakers were arrayed. This laser pointer provided stimuli of 10-ms duration.

Procedure

The whole experiment consisted of two unimodal preexperiments (auditory and visual) and a main experiment comprising an emotional-learning phase and an audiovisual (ventriloquism) phase (see Fig. 1). The order of the unimodal preexperiments was counterbalanced across participants. In the auditory preexperiment, sinusoidal tones were presented randomly from one of the eight loudspeakers, while in the visual preexperiment, the light stimuli were presented at the same set of locations. The participants were instructed to localize the stimuli while fixating the projection of the center-pointing laser. The spatial response was given with a response pointer located in front of the participant, by aligning it accordingly and then pressing a button attached to the pointing device in order to confirm the response and terminate the trial. Subsequently, the next trial started 250 ms after the participant had realigned the pointing device to the midline (±10 °; negative degree values indicate leftward deviation from the midline) and had pressed the button again. Each of the two unimodal experiments consisted of 192 trials (24 stimuli per location).

The data from these preexperiments allowed us to check for any baseline differences between the experimental and control groups in terms of their auditory and visual localization accuracy. Following the two preexperiments, the emotional-learning phase of the study was conducted. The bisyllabic nonsense words "dedu," "none," and "raro" were played randomly and equiprobably from one of the loudspeakers, in a random order. The participants had to localize the sounds with the pointing device in the same manner as described for the preexperiment, all the while fixating the red light presented by means of the central laser beam. The purpose of the emotional-learning phase of the experiment was to associate the auditory stimuli presented from one particular location with an emotional meaning. In order to
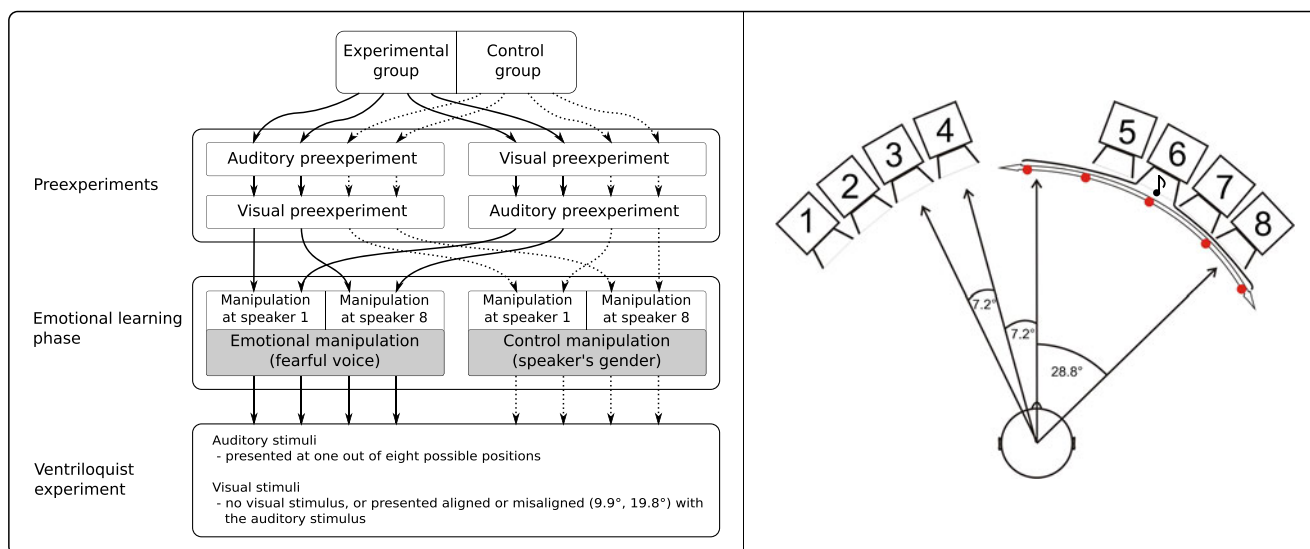
Fig. 1 Study design and experimental setup. The flow diagram shows the different stages of the experiment (preexperiments, emotional-learning phase, and ventriloquist experiment) and how the participants in the study were subdivided to take part in these stages (solid arrows indicate the experimental group; dashed arrows, the control group). The other diagram shows the experimental setup for the audiovisual phase of the main experiment ("ventriloquist experiment"). The participants were seated in the middle of a circular loudspeaker array. Sounds were randomly played from one of the eight loudspeakers and could be presented alone or together with concurrent lights from various locations, relative to the position of the loudspeaker (marked by dots)

control for any nonspecific attentional effects attributable to stimulus salience during the emotional-learning phase of the study, a control group was included. Instead of a voice with an aversive emotional prosody, these participants heard an actor of the gender opposite that of the voices presented from the other seven loudspeakers. Half of the participants were pseudorandomly assigned to the experimental group, the other half to the control group (see Fig. 1). In addition, the groups of participants were divided pseudorandomly (but orthogonally to the experimental/control group split) into two groups of 18 individuals each—the left and right groups. For the left experimental group, the emotionally aversive (fearful) stimuli were presented from the leftmost loudspeaker, whereas for the right experimental group these stimuli were presented from the rightmost loudspeaker instead. The stimuli from all other loudspeakers were spoken in a neutral voice. All of the stimuli presented to the experimental group were spoken by a single female actor. The stimuli presented to the control group were spoken by a male actor if they originated from the leftmost loudspeaker (for the left control group) or from the rightmost loudspeaker (for the right control group). The pseudowords presented at any of the other loudspeakers were spoken by a female actor. All of the stimuli presented to the control group were spoken with an emotionally neutral prosody. The emotional-learning phase consisted of 192 trials (eight trials for each of the combinations of location and word).

The audiovisual phase of the experiment was conducted after the emotional-learning phase. On each trial, short tones were presented. These auditory stimuli were presented either in isolation (unimodal auditory condition) or together with a concurrent visual stimulus at a spatial discrepancy of 0 °, ±9.9 °, or ±19.8 °.

The participants had to try to localize the tones by means of the pointer (as in the preexperiments and the emotional-learning phase of the study). They were explicitly told to ignore the visual stimuli and to fixate the projection of the central laser beam at the midline. An additional task was introduced in order to ensure that the participants did not close their eyes. At random times during the course of the experiment, the fixation point was blinked. After the presentation of the auditory and visual stimuli (10 ms after trial onset), the laser pointer was switched off for 150 ms, then it was alternately switched on and off nine times every 150 ms (i.e., switched off four times and switched on five times). In each block, this deviant trial was presented once (with a probability of .5), twice (with a probability of .25), or not at all (with a probability of .25). The participants had to count the number of changes in the fixation light, without being informed about their probability, and to report after each block of trials. These rare trials were excluded from the subsequent data analysis.[1]

The audiovisual phase of the main experiment consisted of three blocks of 208 trials. In each block of trials, the unimodal auditory condition (i.e., no visual stimulus) and the bimodal conditions with visual stimuli at spatial

---

[1] All of the participants provided the correct number of rare deviants throughout the experiment, with the single exception of one participant, who reported an incorrect number in one of the blocks.

discrepancies of –9.9 °, 0 °, and 9.9 ° with respect to the auditory stimulus were presented six times for each of the loudspeakers (i.e., with each of the loudspeakers as the sound source). The conditions –19.8 ° and 19.8 ° were presented once per block and loudspeaker. All of these conditions (–19.8 °, –9.9 °, 0 °, 9.9 °, 19.8 °, and the unimodal auditory condition) were presented in a random order and at randomly ordered loudspeaker positions in each block.

In the ±19.8 ° conditions, only a weak ventriloquist effect was expected, because audiovisual stimuli with such a large spatial disparity were unlikely to be perceived as originating from a common source (see, e.g., Jackson, 1953). These two conditions were infrequently presented in order to prevent learning effects and the development of idiosyncratic response strategies on the part of the participants. Using only two discrepancy conditions would have carried the risk that participants might have based their responding on the statistics of the stimuli; that is, they might have become aware of the discrepancies and consciously tried to compensate for their presence.

Data analysis Statistical analyses were conducted on the localization data from the preexperiments and the audiovisual phase of the main experiment. The spatial response data (azimuthal response locations) were prepared by subtracting either the spatial positions (in degrees) of the tones (for the auditory preexperiment and the audiovisual phase of the main experiment) or the positions of the visual stimuli (for the visual preexperiment) from the spatial positions of the responses (in degrees) for all trials. As a result, deviation scores (the perceived location minus the veridical location) were derived. These difference scores are termed "deviations" below. Each participant's deviations were averaged for each loudspeaker position in the auditory preexperiment. The deviation scores were averaged for each visual location for the visual preexperiment, and for each combination of loudspeaker position and visual condition for the audiovisual phase of the main experiment.

The data from the left and right groups (see above) were realigned so that the "manipulated" loudspeaker was always the leftmost one. This was done prior to averaging the deviations of the preexperiments and the deviations of the audiovisual phase of the main experiment across participants.

The realignment of the data from the audiovisual phase of the main experiment was performed by flipping the deviations of the right group across the eight loudspeaker positions. Hence, the loudspeaker positions are labeled "1|8," "2|7," "3|6," "4|5," "5|4," 6|3," "7|2," and "8|1" in the following discussion. These collapsed data were multiplied by minus one and additionally flipped across the five bimodal visual conditions per loudspeaker (–19.8 °, –9.9 °, 0 °, 9.9 °, and 19.8 °) in order to preserve the eccentricity of the responses. The resulting data set was used to perform

analyses of variance (ANOVAs) and to compute the means shown in Fig. 3 below.

To test for any possible group differences (emotional-learning vs. control group) regarding the auditory and visual localization performance in the preexperiments, the data from the preexperiments were collapsed in a manner analogous to that described above prior to calculating the group means. The response data from the right group were flipped across the eight loudspeaker positions so that the "manipulated" loudspeaker was always the leftmost position, and the data points were multiplied by minus one in order to preserve the eccentricity of the responses. For the auditory and visual preexperiments, as well as for the audiovisual phase of the main experiment, these collapsed data sets will be referred to as the "realigned data sets" below.

Since for loudspeaker positions 4|5 and 5|4 the auditory and visual stimuli were actually presented from different hemifields in the audiovisual phase of the main experiment, these positions were disregarded in order to avoid any possible meridian effects (cf. Rorden & Driver, 2001; Tassinari & Campara, 1996). Data from the visual conditions –19.8 ° and +19.8 ° were also excluded from the data analysis because the trial numbers for these conditions were much lower and, as expected, the ventriloquist effect was markedly reduced for these audiovisual disparities.

Results

Preexperiments

The deviations are depicted in Fig. 2a and b for the auditory and visual preexperiments, respectively. These figures highlight the realigned data sets, meaning that the data had been collapsed so that the critical loudspeaker was always the outermost left one. An ANOVA with the factors Hemifield (left and right), Eccentricity (1, 2, 3, and 4), Modality (auditory and visual), and Group (experimental and control) was conducted on the combined response data from the two preexperiments. The analysis yielded a significant hemifield effect [$F(1, 34) = 5.19$, $p = .03$], indicating that the eccentricities of the stimuli had been overestimated in the preexperiments.[2]

Neither the Group factor [$F(1, 34) = 1.07$, $p = .31$] nor the Group × Modality interaction [$F(1, 34) = 0.17$, $p = .68$], nor

---

[2] The overestimation of the eccentricities of the stimuli is expressed by the fact that the deviations are negative for the "left" four loudspeaker positions and positive for the remaining four positions (see Fig. 2). Hence, the main effect of hemifield confirms the fact that the eccentricity of the stimuli was overestimated by participants. A significant effect of eccentricity would have indicated that the amount of overestimation varied across loudspeakers at different eccentricities.

Fig. 2 Mean deviations of the spatial responses from the actual sound location are depicted for all eight loudspeaker positions in the preexperiments. Before we calculated group averages, the data were realigned so that loudspeaker position 1|8 was the critical position in the emotional-learning phase of the experiment. (a) Results from the auditory preexperiment. (b) Results from the visual preexperiment. Negative values indicate positions to the left of the midline, and positive ones indicate positions to the right of midline. Error bars denote standard errors
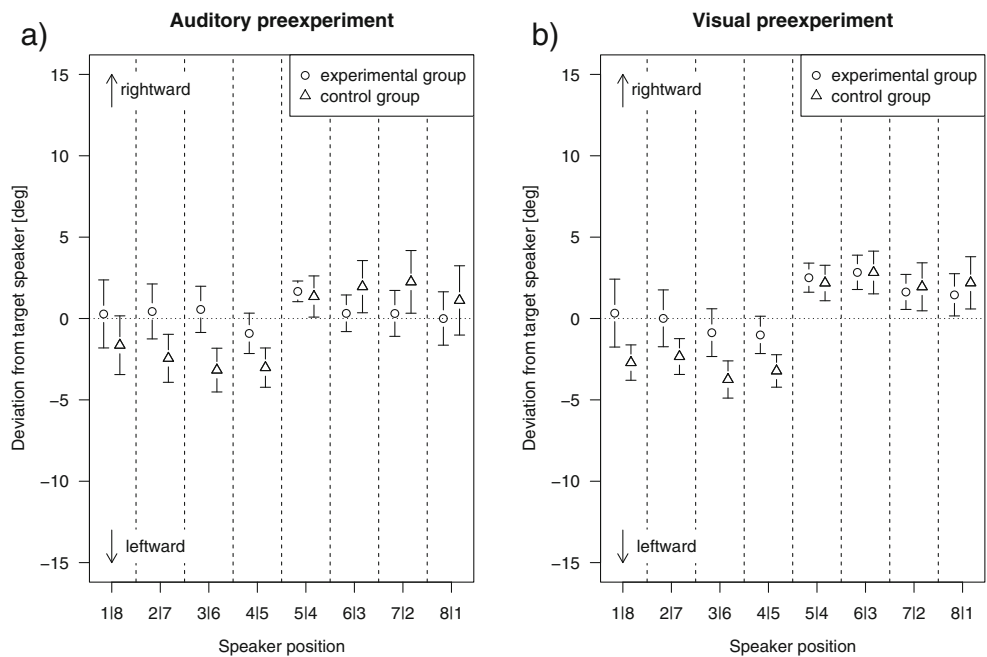
any higher-order interaction with the Group factor, reached statistical significance. This means that visual and auditory localization accuracies were comparable between the two groups before the emotional-learning phase of the study.[3]

Audiovisual phase of the main experiment

Figure 3 shows the spatial responses of the experimental and control groups in the audiovisual phase of the main experiment. The deviations for the five possible visual–auditory combinations for each of the eight loudspeakers are plotted. In addition, the responses to unimodal stimuli are depicted (see condition "a").
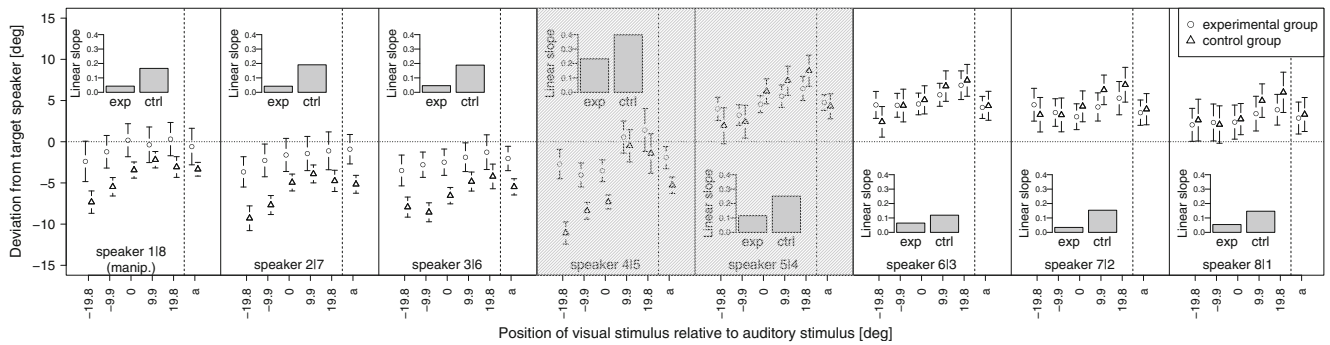
Bimodal conditions In general, it can be seen in Fig. 3 that the eccentricity of the stimuli was overestimated by participants. More importantly, a ventriloquist effect was induced at all loudspeaker positions; that is, the deviations followed the positions of the visual stimulus. This is highlighted by a positive slope of the deviations across the visual conditions –19.8 ° to +19.8 ° for each loudspeaker (see the inset bar plots

[3] The main effect of group was not significant, indicating that the experimental group and the control group did not differ in their pre-experimental localization performance. However, it appears from Fig. 2 that there might be group differences at individual speaker locations. When running separate *t* tests at every loudspeaker location (experimental group vs. control group), a marginal group difference was observed at loudspeaker position 3|6 (experimental group, $M = 0.56$, $SE = 1.42$; control group, $M = –3.17$, $SE = 1.34$) [$t(33.90) = –1.91$, $p = .06$] that did not survive a Šidák correction for multiple tests in the auditory preexperiment, and no (marginally) significant group differences in the visual preexperiment.

in Fig. 3), indicating that the auditory localization responses were biased toward the respective positions of the visual stimuli. This effect was larger for visual locations –9.9 ° to +9.9 ° than for visual locations –19.8 ° and +19.8 °. The latter locations are not considered any further (see the Data Analysis section above). To test the hypothesis that the ventriloquist effect was stronger in the control than in the experimental group, an ANOVA was conducted with the factors Hemifield (left and right), Visual Location (–9.9 °, 0 °, and 9.9 °), and Speaker Eccentricity (2, 3, and 4), as well as the between-subjects factor Group (experimental and control). The ANOVA yielded significant effects for the factors Hemifield [$F(1, 34) = 16.36$, $p < .001$] and Visual Location [$F(2, 68) = 13.65$, $p < .001$], as well as significant interactions of Hemifield × Visual Location [$F(2, 68) = 3.15$, $p < .05$], Hemifield × Eccentricity [$F(2, 68) = 17.94$, $p < .001$], and Group × Visual Location [$F(2, 68) = 4.05$, $p = .04$].

Since the visual-location effect reflects a bias of auditory localization toward the location of the visual stimulus—that is, a ventriloquist effect—the Group × Visual Location interaction indicates a reduced ventriloquist effect in the experimental as compared to the control group.

Apart from the ventriloquist effect seen in the data, it would appear from Fig. 3 that in the hemifield in which the critical loudspeaker was positioned, the eccentricity of the auditory stimuli was generally overestimated across all five visual locations by the control group. By contrast, the responses of the experimental group appear to be closer to the actual location of the stimulus, meaning that the eccentricity effect (as expressed by the significant effect for the Hemifield factor) was less pronounced for this group in the

Fig. 3 Results from the audiovisual ventriloquism experiment. For both groups of participants (experimental and control), mean deviations from the true sound source are shown per loudspeaker position and visual condition. The data are aligned so that Position 1|8 is the location that had been manipulated in the emotional-learning phase. Data from the unimodal auditory condition (i.e., no visual stimulus) are designated by "a" in each panel. Error bars denote standard errors. The inset bar plots show the slopes of linear regression lines fitted to the mean deviations across the visual positions –9.9 °, 0 °, and 9.9 ° for the two groups at each loudspeaker location. These slopes reflect the ventriloquist effect. The data from the shaded loudspeaker positions were discarded from the statistical analysis, due to possible meridian effects

hemifield with the critical manipulation. However, the ANOVA did not reveal a Group × Hemifield interaction [$F(1, 34) = 1.43$, $p = .24$], nor any related higher-order interactions including the Group factor.[4]

Unimodal auditory condition An ANOVA with the factors Hemifield (left and right), Eccentricity (2, 3, and 4), and Group (experimental and control) was conducted on the realigned deviations from the unimodal auditory condition. As a result, a significant main effect of the Hemifield factor [$F(1, 34) = 14.03$, $p < .001$] and a Hemifield × Eccentricity interaction [$F(2, 68) = 5.29$, $p = .02$] were observed. This result indicates that the eccentricity of the auditory stimuli was overestimated by participants, and that the influence of the eccentricity of the stimuli on the responses varied across the two hemifields. No group effect [$F(1, 34) = 1.39$, $p = .25$] nor any interaction including the Group factor was observed.[5]

Comparison of the preexperimental data
with the audiovisual phase of the main experiment

To test for group-specific changes of auditory localization accuracy due to the emotional-learning phase, the combined realigned data sets from the auditory preexperiment and the unimodal condition of the audiovisual phase of the main experiment were subjected to an ANOVA with the factors Hemifield (left and right), Eccentricity (2, 3, and 4), and Experiment (pre and main), as well as the between-subjects factor Group (experimental and control). This analysis revealed a significant main effect of hemifield [$F(1, 34) = 6.74$, $p < .05$], a marginally significant Eccentricity × Hemifield interaction [$F(2, 68) = 3.73$, $p = .05$], and a significant Hemifield × Experiment interaction [$F(1, 34) = 7.01$, $p < .05$], but no significant Group × Experiment interaction [$F(1, 34) = 1.16$, $p = .29$], nor any significant higher-order interaction including the Group factor. This result indicates that no group-specific change in auditory localization accuracy occurred due to the emotional-learning phase.[6]

Discussion

The present study was conducted to determine whether crossmodal binding, as assessed by the ventriloquist effect,

---

[4] Nevertheless, we ran separate *t* tests at the loudspeaker positions 1|8, 2|7, and 3|6 between the responses of the two groups. When pooling the data from the three visual conditions (–9.9 °, 0 °, and 9.9 °), we found a significant group difference at the loudspeaker position 3|6 (experimental group, $M = –2.38$, $SE = 1.55$; control group, $M = –6.64$, $SE = 1.06$) [$t(30) = –2.27$, $p = .03$], but not at positions 1|8 and 2|7. When we tested the 0 ° condition separately, we also found a significant group difference at loudspeaker position 3|6 (experimental group, $M = –2.48$, $SE = 1.60$; control group, $M = –6.54$, $SE = 0.99$) [$t(28.46) = –2.16$, $p = .04$], but not at the other two positions. These results were not significant after Šidák correction for three tests.

[5] Subsequent *t* tests at loudspeaker positions 1|8, 2|7, and 3|6 revealed a marginal group difference at loudspeaker positions 2|7 (experimental group, $M = –0.91$, $SE = 1.79$; control group, $M = –5.16$, $SE = 1.08$) [$t(28.02) = –2.04$, $p = .05$] and 3|6 (experimental group, $M = –2.03$, $SE = 1.48$; control group, $M = –5.47$, $SE = 0.98$) [$t(29.59) = –1.94$, $p = .06$]. These results were not marginally significant after Šidák correction for three tests.

[6] When the responses of the auditory preexperiment were compared to the 0 ° condition of the audiovisual phase of the main experiment, a similar pattern of results was found—that is, neither the Group × Experiment interaction [$F(1, 34) = 0.89$, $p = .35$] nor any higher-order interactions including the Group factor reached statistical significance. In essence, this result indicates that no group-specific modulation of the bimodal enhancement (as compared to the preexperimental "baseline") in localization accuracy was observed due to the emotional-learning phase.

is an automatic process or rather is influenced by the emotional systems. An emotional-learning paradigm was used in which bisyllabic pseudowords were presented from one of eight possible loudspeaker locations, four situated in either hemifield. The emotional prosody (experimental group) of the pseudowords heard from the outermost loudspeaker on either the left or the right was manipulated: The words were spoken in a fearful instead of a neutral prosody. In each trial of the subsequent audiovisual experiment, a pure tone was presented from one of the eight loudspeakers, together with a visual stimulus at the same position or at a location to the left or right of each loudspeaker, or without a concurrent visual stimulus (unimodal auditory condition). The participants were instructed to localize the auditory stimulus while trying to ignore the accompanying visual stimulus.

The analysis of the data from the audiovisual experiment revealed that emotional learning reduced the magnitude of the subsequently measured ventriloquist effect. This effect was not spatially selective; that is, it was observed at all loudspeaker locations. To the best of our knowledge, this is the first demonstration that audiovisual binding can be modulated by emotional learning. In order to answer the question whether this result argues against the full automaticity of the audiovisual binding process as assessed by the ventriloquist illusion, different mechanisms that could potentially explain our findings have to be taken into account.

First, emotional learning might have resulted in enhanced processing of the auditory stimuli (Anderson & Phelps, 2001; Phelps, Ling, & Carrasco, 2006; Vuilleumier, 2005). This might have been mediated by, for example, a more precise tuning of auditory representations ("auditory sharpening"; see Weinberger, 2004). According to the maximum-likelihood account of crossmodal integration (Alais & Burr, 2004; Ernst & Bülthoff, 2004), any increase in the reliability of the auditory input should increase the weight assigned to that input, and thus reduce the magnitude of the ventriloquist effect. However, since we did not find a significant improvement in auditory localization performance in the experimental as compared to the control group, this hypothesis was not confirmed by the empirical data reported here. In particular, group differences were found neither in the unimodal auditory condition of the main experiment, nor in a comparison between preexperimental performance and the unimodal auditory condition, which tested for changes in auditory localization performance due to the emotional-learning phase.

Alternatively, the process of audiovisual binding might have changed as a consequence of emotional learning. More specifically, the aversive stimulus presented in the auditory modality might have influenced processes of intermodal selection, in favor of audition. From the perspective of cognitive control (see Pessoa, 2009), the ventriloquist illusion constitutes a typical conflict situation in which distractor processing has to be inhibited. Cognitive control

functions are associated with the prefrontal cortex (PFC; Botvinick, Cohen, & Carter, 2004; Meienbrock, Naumer, Doehrmann, Singer, & Muckli, 2007; Miller, 2000). Interestingly, the PFC (particularly the orbitofrontal cortex, OFC) has been considered an important brain structure for crossmodal processing as well (Chavis & Pandya, 1976; Jones & Powell, 1970; Kringelbach, 2005). Thus, emotional learning might be expected to affect intermodal selection by inhibiting visual distractor processing more strongly. This gives rise to the question of how PFC activity might be affected by emotional conditioning. Emotionally meaningful stimuli, including voices, are processed by the amygdala (Fecteau, Belin, Joanette, & Armony, 2007; Frühholz, Ceravolo, & Grandjean, 2012; Klinge et al., 2010; Sander & Scheich, 2001; Wildgruber, Ackermann, Kreifelts, & Ethofer, 2006). Animal studies on fear conditioning (Bordi & LeDoux, 1992; Phillips & LeDoux, 1992; Quirk, Armony, & LeDoux, 1997) and neuroimaging studies in humans (Büchel, Morris, Dolan, & Friston, 1998) have provided evidence that the amygdala plays a pivotal role, especially in the acquisition phase of emotional learning (see also Damasio et al., 2000; Dougherty et al., 1999). What is more, the amygdala is densely connected to the PFC, and especially to the OFC (Cavada, Compañy, Tejedor, Cruz-Rizzolo, & Reinoso-Suárez, 2000; Salzman & Fusi, 2010; Vuilleumier, 2005). Given the central role that the amygdala plays in the acquisition of the emotional valence of stimuli, as well as the involvement of the OFC in crossmodal integration and control processes, it might be speculated that these two brain regions are involved in a neural system that detects the emotional significance of stimuli and, consequently, adjusts the level of cognitive control that is associated with the intermodal selection of the stimuli. This idea is further supported by studies reporting the involvement of the OFC in the processing of emotionally meaningful stimuli (Rolls, 1999; Royet et al., 2000)—in particular, of aversive prosody (Paulmann, Seifert, & Kotz, 2010; Wildgruber et al., 2006). Regarding the present results, the interplay between the amygdala and (orbito)frontal cortex in response to the presentation of fearful stimuli might have increased cognitive control in such a way that intermodal interference by the visual distractor was diminished, thus resulting in a reduction of the ventriloquist illusion in the experimental group.

This account would also be compatible with (1) our result that unimodal auditory localization performance did not change as a consequence of emotional learning and (2) the lack of spatial selectivity of the reduction of the ventriloquist effect that we measured in the experimental group.

In addition, this explanation of the data is compatible with previous results on the Colavita visual dominance effect: When performing an audiovisual detection task, humans tend to preferentially report the visual part of

audiovisual stimuli, reflecting overall visual dominance (Colavita, 1974; see also Spence, 2009, for a review). Interestingly, Shapiro, Egerman, and Klein (1984) combined a typical Colavita experiment with the unpredicted presentation of aversive electric shocks. Under such conditions, a reversal of the Colavita effect was observed. That is, the auditory part of the stimulus was reported more often in the experimental group (with unpredictable shock) than in the control group (without shock). Hence, just as in our study, modality selection was readjusted as a result of emotional learning. Shapiro et al. argued that the Colavita effect was reversed in their experiment due to a general bias toward the auditory stimulus when threatening stimuli are expected. Auditory stimuli can be detected from all locations around the body, including those outside the current field of view. Consequently, the biological relevance of auditory stimuli is increased in threatening situations, which might provide an explanation for the dominance of audition following threatening stimuli. In the present study, the threatening stimulus was presented in the auditory modality. Readjusting modality dominance in this situation constitutes a plausible mechanism, given that the system might expect more threatening stimuli—but possibly not only at the same location where such stimuli had been perceived previously.

More generally, when Shapiro et al.'s (1984) results are taken together with the present results, the suggestion that emerges is that crossmodal binding is not automatic and exclusively determined by the reliability of the available input modalities (Ernst & Bülthoff, 2004). Rather, it would appear that the emotional significance of the stimuli presented, and the resulting relevance of each modality channel, need to be considered as additional determinants of the respective weights given to each modality in crossmodal integration. In summary, therefore, it seems justified to conclude that crossmodal binding, as assessed by the ventriloquist effect, is not fully automatic. Whether the observed modulation of crossmodal binding is specific to the case of emotional relevance—that is, to interactions with the emotional system—or also holds true for other influences, such as from the reward system, should certainly be tested in future research.

References

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology, 14*, 257–262.

Anderson, A. K. (2005). Affective influences on the attentional dynamics supporting awareness. *Journal of Experimental Psychology. General, 134*, 258–281. doi:10.1037/0096-3445.134.2.258

Anderson, A. K., & Phelps, E. A. (2001). Lesions of the human amygdala impair enhanced perception of emotionally salient events. *Nature, 411*, 305–309.

Bakin, J. S., & Weinberger, N. M. (1990). Classical conditioning induces CS-specific receptive field plasticity in the auditory cortex of the guinea pig. *Brain Research, 536*, 271–286.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review, 5*, 482–489. doi:10.3758/BF03208826

Bertelson, P., & de Gelder, B. (2004). The psychology of multimodal perception. In C. Spence & J. Driver (Eds.), *Crossmodal space and crossmodal attention* (pp. 141–177). Oxford, U.K.: Oxford University Press.

Bertelson, P., Pavani, F., Ladavas, E., Vroomen, J., & de Gelder, B. (2000a). Ventriloquism in patients with unilateral visual neglect. *Neuropsychologia, 38*, 1634–1642. doi:10.1016/S0028-3932(00)00067-1

Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000b). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics, 62*, 321–332. doi:10.3758/BF03205552

Bonath, B., Noesselt, T., Martinez, A., Mishra, J., Schwiecker, K., Heinze, H.-J., & Hillyard, S. A. (2007). Neural basis of the ventriloquist illusion. *Current Biology, 17*, 1697–1703.

Bordi, F., & LeDoux, J. (1992). Sensory tuning beyond the sensory system: An initial analysis of auditory response properties of neurons in the lateral amygdaloid nucleus and overlying areas of the striatum. *Journal of Neuroscience, 12*, 2493–2503.

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8*, 539–546. doi:10.1016/j.tics.2004.10.003

Büchel, C., Morris, J., Dolan, R. J., & Friston, K. J. (1998). Brain systems mediating aversive conditioning: An event-related fMRI study. *Neuron, 20*, 947–957.

Cavada, C., Compañy, T., Tejedor, J., Cruz-Rizzolo, R. J., & Reinoso-Suárez, F. (2000). The anatomical connections of the macaque monkey orbitofrontal cortex. A review. *Cerebral Cortex, 10*, 220–242.

Chavis, D. A., & Pandya, D. N. (1976). Further observations on corticofrontal connections in the rhesus monkey. *Brain Research, 117*, 369–386.

Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics, 16*, 409–412.

Damasio, A. R., Grabowski, T. J., Bechara, A., Damasio, H., Ponto, L. L. B., Parvizi, J., & Hichwa, R. D. (2000). Subcortical and cortical brain activity during the feeling of self-generated emotions. *Nature Neuroscience, 3*, 1049–1056.

Dougherty, D. D., Shin, L. M., Alpert, N. M., Pitman, R. K., Orr, S. P., Lasko, M., . . . Rauch, S. L. (1999). Anger in healthy men: A PET study using script-driven imagery. *Biological Psychiatry, 46*, 466–472.

Driver, J. (1996). Enhancement of selective listening by illusory mislocation of speech sounds due to lip-reading. *Nature, 381*, 66–68.

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences, 8*, 162–169.

Fairhall, S. L., & Macaluso, E. (2009). Spatial attention can modulate audiovisual integration at multiple cortical and subcortical sites. *European Journal of Neuroscience, 29*, 1247–1257. doi:10.1111/j.1460-9568.2009.06688.x

Fecteau, S., Belin, P., Joanette, Y., & Armony, J. L. (2007). Amygdala responses to nonlinguistic emotional vocalizations. *NeuroImage, 36*, 480–487.

Fox, E. (2002). Processing emotional facial expressions: The role of anxiety and awareness. *Cognitive, Affective, & Behavioral Neuroscience, 2*, 52–63. doi:10.3758/CABN.2.1.52

Fox, E., Russo, R., & Dutton, K. (2002). Attentional bias for threat: Evidence for delayed disengagement from emotional faces. *Cognition & Emotion, 16*, 355–379.

Frühholz, S., Ceravolo, L., & Grandjean, D. (2012). Specific brain networks during explicit and implicit decoding of emotional prosody. *Cerebral Cortex, 22*(5), 1107–1117. doi:10.1093/cercor/bhr184

Jackson, C. V. (1953). Visual factors in auditory localization. *Quarterly Journal of Experimental Psychology, 5*, 52–65.

Jones, E. G., & Powell, T. P. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain, 93*, 793–820.

Kitagawa, N., & Spence, C. (2005). Investigating the effect of a transparent barrier on the crossmodal congruency effect. *Experimental Brain Research, 161*, 62–71.

Klinge, C., Röder, B., & Büchel, C. (2010). Increased amygdala activation to emotional auditory stimuli in the blind. *Brain, 133*, 1729–1736.

Kringelbach, M. L. (2005). The human orbitofrontal cortex: Linking reward to hedonic experience. *Nature Reviews Neuroscience, 6*, 691–702.

Meienbrock, A., Naumer, M. J., Doehrmann, O., Singer, W., & Muckli, L. (2007). Retinotopic effects during spatial audio-visual integration. *Neuropsychologia, 45*, 531–539.

Miller, E. K. (2000). The prefrontal cortex and cognitive control. *Nature Reviews Neuroscience, 1*, 59–65.

Paulmann, S., Seifert, S., & Kotz, S. A. (2010). Orbito-frontal lesions cause impairment during late but not early emotional prosodic processing. *Social Neuroscience, 5*, 59–75. doi:10.1080/17470910903135668

Pessoa, L. (2009). How do emotion and motivation direct executive control? *Trends in Cognitive Sciences, 13*, 160–166.

Phelps, E. A., Ling, S., & Carrasco, M. (2006). Emotion facilitates perception and potentiates the perceptual benefits of attention. *Psychological Science, 17*, 292–299.

Phillips, R. G., & LeDoux, J. E. (1992). Differential contribution of amygdala and hippocampus to cued and contextual fear conditioning. *Behavioral Neuroscience, 106*, 274–285. doi:10.1037/0735-7044.106.2.274

Quirk, G. J., Armony, J. L., & LeDoux, J. E. (1997). Fear conditioning enhances different temporal components of tone-evoked spike trains in auditory cortex and lateral amygdala. *Neuron, 19*, 613–624.

Radeau, M., & Bertelson, P. (1974). The after-effects of ventriloquism. *Quarterly Journal of Experimental Psychology, 26*, 63–71.

Röder, B., & Büchel, C. (2009). Multisensory interactions within and outside the focus of visual spatial attention (commentary on Fairhall & Macaluso). *European Journal of Neuroscience, 29*, 1245–1246.

Rolls, E. T. (1999). *The brain and emotion*. New York, NY: Oxford University Press.

Rorden, C., & Driver, J. (2001). Spatial deployment of attention within and across hemifields in an auditory task. *Experimental Brain Research, 137*, 487–496.

Royet, J.-P., Zald, D., Versace, R., Costes, N., Lavenne, F., Koenig, O., & Gervais, R. (2000). Emotional responses to pleasant and unpleasant olfactory, visual, and auditory stimuli: A positron emission tomography study. *Journal of Neuroscience, 20*, 7752–7759.

Salzman, C. D., & Fusi, S. (2010). Emotion, cognition, and mental state representation in amygdala and prefrontal cortex. *Annual Review of Neuroscience, 33*, 173–202.

Sander, K., & Scheich, H. (2001). Auditory perception of laughing and crying activates human amygdala regardless of attentional state. *Cognitive Brain Research, 12*, 181–198.

Shapiro, K. L., Egerman, B., & Klein, R. M. (1984). Effects of arousal on human visual dominance. *Perception & Psychophysics, 35*, 547–552.

Spence, C. (2009). Explaining the Colavita visual dominance effect. *Progress in Brain Research, 176*, 245–258.

Talsma, D., Senkowski, D., Soto-Faraco, S., & Woldorff, M. G. (2010). The multifaceted interplay between attention and multisensory integration. *Trends in Cognitive Sciences, 14*, 400–410. doi:10.1016/j.tics.2010.06.008

Tassinari, G., & Campara, D. (1996). Consequences of covert orienting to non-informative stimuli of different modalities: A unitary mechanism? *Neuropsychologia, 34*, 235–245. doi:10.1016/0028-3932(95)00085-2

Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics, 63*, 651–659. doi:10.3758/BF03194427

Vroomen, J., & de Gelder, B. (2004). Perceptual effects of cross-modal stimulation: Ventriloquism and the freezing phenomenon. In G. A. Calvert, C. Spence, & B. E. Stein (Eds.), *The handbook of multisensory processes* (pp. 141–150). Cambridge, MA: MIT Press.

Vuilleumier, P. (2005). How brains beware: Neural mechanisms of emotional attention. *Trends in Cognitive Sciences, 9*, 585–594.

Weinberger, N. M. (2004). Specific long-term memory traces in primary auditory cortex. *Nature Reviews Neuroscience, 5*, 279–290.

Welch, R. B. (1999). Meaning, attention, and the unity assumption in the intersensory bias of spatial and temporal perceptions. In G. Aschersleben, T. Bachmann, & J. Musseler (Eds.), *Cognitive contributions to perception of spatial and temporal events* (pp. 371–387). New York, NY: Elsevier.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin, 88*, 638–667.

Wildgruber, D., Ackermann, H., Kreifelts, B., & Ethofer, T. (2006). Cerebral processing of linguistic and emotional prosody: fMRI studies. *Progress in Brain Research, 156*, 249–268.

Williams, J. M. G., Mathews, A., & MacLeod, C. (1996). The emotional Stroop task and psychopathology. *Psychological Bulletin, 120*, 3–24. doi:10.1037/0033-2909.120.1.3