Multisensory aversive stimuli differentially modulate negative feelings in near and far space

Marine Taffou, Jan Ondřej, Carol O 'Sullivan, Olivier Warusfel, Stéphanie Dubal, Isabelle Viaud-Delmon

Multisensory aversive stimuli differentially modulate negative feelings in near and far space

Marine Taffou[1, 2*], Jan Ondřej[3], Carol O'Sullivan[3], Olivier Warusfel[1], Stéphanie Dubal[2], Isabelle Viaud-Delmon[1]

[1]Sciences et Technologies de la Musique et du Son, CNRS UMR 9912, IRCAM, Sorbonne Universités, UPMC Univ Paris 06, 1 place Igor Stravinsky, F-75004 Paris, France

[2]Inserm, U 1127, CNRS UMR 7225, Sorbonne Universités, UPMC Univ Paris 06 UMR S 1127, Institut du Cerveau et de la Moelle épinière, ICM, Social and Affective Neuroscience (SAN) Laboratory, 47 boulevard de l'hôpital, F-75013 Paris, France

[3]School of Computer Science and Statistics, Trinity College Dublin, Dublin 2, Ireland

*Corresponding author.

Marine Taffou
1 place Igor Stravinsky
75004 Paris – France
Mail: marine.taffou@ircam.fr
Tel: +33 (0)1 44 78 15 55
Fax: +33 (0)1 44 78 15 45

Abstract

Affect, space and multisensory integration are processes that are closely linked. However, it is unclear whether the spatial location of emotional stimuli interacts with multisensory presentation to influence the emotional experience they induce in the perceiver.

In this study, we used the unique advantages of virtual reality techniques to present potentially aversive crowd stimuli embedded in a natural context and to control their display in terms of sensory and spatial presentation. Individuals high in crowdphobic fear navigated in an auditory-visual virtual environment, in which they encountered virtual crowds presented through the visual channel, the auditory channel or both. They reported the intensity of their negative emotional experience at a far distance and at a close distance from the crowd stimuli.

Whereas auditory-visual presentation of close feared stimuli amplified negative feelings, auditory-visual presentation of distant feared stimuli did not amplify negative feelings.

This suggests that spatial closeness allows multisensory processes to modulate the intensity of the emotional experience induced by aversive stimuli. Nevertheless, the specific role of auditory stimulation must be investigated in order to better understand this interaction between multisensory, affective and spatial representation processes. This phenomenon may serve the implementation of defensive behaviors in response to aversive stimuli that are in position to threaten an individual's feeling of security.

Keywords: auditory-visual integration; emotion; spatial cognition; peripersonal space; crowd; virtual reality

1. Introduction

In daily life, affective events often convey emotional cues through multiple senses. For example, the joy of a crowd of people celebrating a win after a sport competition is perceived via vision (the crowd jumps and dances) and via audition (the crowd laughs and sings). While the processing of affective events has primarily been studied in one sensory modality at a time – mostly vision –, the investigation of how we deal with multisensory affective events has emerged over the last twenty years. This emergence was spurred on by the explosion of research activity on multisensory processing, which has indicated that the multiplicity of sensory cues about a neutral event leads to a faster detection (Diederich & Colonius, 1987; Giray & Ulrich, 1993; Gondan, Lange, Rösler, & Röder, 2004; Miller, 1982), a faster recognition (Gondan, Niederhaus, Rösler, & Röder, 2005; Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Miller, 1991; Molholm, Ritter, Javitt, & Foxe, 2004; Suied, Bonneel, & Viaud-Delmon, 2009) and a more accurate comprehension (Risberg & Lubker, 1978; Sumby & Pollack, 1954) of the event and that these perceptual and behavioural gains are, at least in part, linked to the combination and integration of sensory information. If combining different sensory information about the surrounding environment leads to a better apprehension of external events, it could be extremely beneficial insofar as affective events are concerned.

Studies investigating the processing of multisensory affective events have mainly used natural auditory-visual events composed of human faces or bodies paired with human voices presented in the form of pictures, sound recordings and video (e.g. Hagan et al., 2013; Kokinous et al., 2015; Watson et al., 2014 – see also Klasen et al., 2012 for a review and Taffou et al., 2013). Similarly to neutral cues, the multiple sensory affective cues seems to be combined and integrated to lead to an identification of the events' emotional significance that is faster (Collignon et al., 2008; De Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999; De Gelder & Vroomen, 2000; Dolan, Morris, & De Gelder, 2001; Föcker, Gondan, & Röder, 2011; Li et al., 2013; Pourtois, De Gelder, Bol, & Crommelinck, 2005) and more accurate (Collignon et al., 2008; Föcker et al., 2011; Kreifelts et al., 2007; Li et al., 2013; Tanaka et al., 2010; Van den Stock, Grèzes, & De Gelder, 2008).

Yet, these studies have mainly focused on the first stages of affective processing. The processing of an affective stimulus consists of several stages from emotion perception with the identification of the emotional significance of the stimulus to emotion induction with the production of an affective state and its regulation (Damasio, 1998; Phillips, Drevets, Rauch, & Lane, 2003). The production of an affective state comprises the production of emotional responses such as autonomic responses as well as the production of a conscious emotional experience, also called feelings. Little is known about the effect of multisensory affective stimuli on the conscious emotional experience induced in the perceiver. A few studies have addressed this question and showed that the intensity of the emotional experience induced by affective events is increased when they convey emotional cues *via* both vision and audition. This has been demonstrated by arbitrarily coupling auditory and visual events composed of affective pictures and music excerpts (Baumgartner, Lutz, Schmidt, & Jäncke, 2006) and with natural multisensory aesthetic events such as musical performances (Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011; Vines, Krumhansl, Wanderley, & Levitin, 2006) and negatively-valenced dance performances (Christensen, Gaigg, Gomila, Oke, & Calvo-Merino, 2014). We also showed that the emotional experience induced by dogs in subjects high in cynophobic fear was increased when they could both see and hear the dogs (Taffou et al., 2013). The present study aims at furthering the understanding of the effect of

multisensory stimuli on conscious emotional experience by taking into account the distance between the perceiver and the emotional stimulus and using natural multisensory events presented in an ecological context.

The influence of multisensory affective events on emotional experience is indeed intermingled with the effect of their spatial distance from the subject. Spatial distance and fear are inextricably linked because close events represent more of a threat than events located farther away (Mobbs et al., 2007). Moreover, it is well established that space is not a unitary construct (Hall, 1966; Hayduk, 1978) and that events located at close or far distance from the body are represented differently in the brain (e.g. Cowey, Small, & Ellis, 1994; Halligan & Marshall, 1991; Previc, 1998; Rizzolatti, Fadiga, Fogassi, & Gallese, 1997). Specifically, the space close to the body is called near (or peri-personal) space. The representation of this space is multisensory by definition and is thought to have different roles: a role in planning approaching motor acts toward a reachable target (see Brozzoli, Makin, Cardinali, Holmes, & Farnè, 2012) but also a social role by determining the quality and quantity of sensory information exchange during social interaction (Aiello, 1987; Hall, 1966) and a protective role by implementing a safety margin around the body, enabling the preparation of defensive behaviors (Dosey & Meisels, 1969; Graziano & Cooke, 2006).

Consequently, it is not clear how the effects of multisensory presentation and closeness of affective events are linked to produce the emotional experience. Does the multisensory presentation of affective stimuli differentially influence the emotional experience induced in the perceiver as a function of the location of the stimulus at close or far distances from the perceiver? Given that fear and anxiety skew our spatial perception and expand the space, within which fearful events are represented as close (Ferri, Tajadura-Jiménez, Väljamäe, Vastano, & Costantini, 2015; Lourenco, Longo, & Pathman, 2011; Sambo & Iannetti, 2013; Taffou & Viaud-Delmon, 2014; Vagnoni, Lourenco, & Longo, 2012), it seemed to us rather clear that affective, multisensory and space representation processes interact. Thus, we expected that distance would play a role in the multisensory modulation of negative emotional experience and that multisensory aversive stimuli would enhance emotional experience when presented at close but not at far distances from the perceiver.

The elucidation of this questions necessitates the use of an experimental setup where the display of stimulation allows participants to evaluate distances with affective events in an egocentric manner and thus to share the same space as the events. The present study used virtual reality (VR) to manipulate the sensory presentation and the spatial location of crowds within a naturalistic virtual environment. Crowds are genuinely aversive for a subset of individuals sensitive to the fear of crowds. Furthermore, crowds of people can convey information via both the auditory and the visual pathways. Besides, the emotional responses to crowds of people are closely linked to the peri-personal space representation and its role in social and defensive behaviors. Finally, little is known on the fear induced by crowds while it is a symptom found in several disorders including panic disorder with agoraphobia, fear of falling and Parkinson disease. Thus, we recruited a non-clinical sample of participants sensitive to the fear of crowds (crowd-fearful [CF] group) to examine emotional experience in response to auditory and visual crowd stimuli. For these participants, crowds are emotional aversive stimuli. We also recruited a non-clinical sample of participants non-sensitive to the fear of crowds (non-fearful [NF] group) as a control group. For these participants, crowds are not aversive stimuli. The sensitivity to the fear of crowds was used to manipulate the attribution of a negative or neutral valence to the crowd stimuli. Participants explored an auditory-visual virtual scene, in which they encountered virtual crowd stimuli presented through the auditory

channel, the visual channel or both channels and were asked to report the intensity of their negative experience when the crowds were located far from them (8m) or close to them (2m).

2. Methods

2.1. Participants

Participants were recruited from the students of the University Pierre et Marie Curie in Paris. They were selected on the basis of their scores on a questionnaire exploring the fear of crowds (https://www.researchgate.net/publication/286780388_Fear_of_crowds_questionnaire). This questionnaire comprises 15 items describing common crowded situations and addressing different aspects of an encounter with a crowd such as the sensory modalities through which the crowd is sensed (auditory, visual, and/or tactile stimulation) or the mobility of the crowd (fixed in space as when "standing in a crowded bus" or in movement as when "walking in a subway station during rush hour"). Each item has to be rated according to the intensity of discomfort they elicit on a scale from 0 (no discomfort) to 3 (extreme discomfort). The minimal total score on this crowd fear questionnaire is 0, with a maximum of 45. Two hundred and twenty-eight individuals (121 women, age: 24.55 ± 5.32) completed this questionnaire. A distribution of scores was obtained and served as a basis to select participants to the current experiment.

Twenty-two healthy individuals with normal audition and vision were recruited to participate in the study. Ten individuals had a low crowd fear score (score < 20th centile) and composed the NF group. The remaining twelve individuals had high crowd fear scores (score > 80th centile) and composed the CF group. Table 1 lists the characteristics for each group of subjects. Gender and age distributions were similar in the NF and CF groups. The responses to self-report psychometric inventories and questionnaires revealed differences among the groups that were consistent with their group definitions.

All participants provided written informed consent prior to the experiment, which was approved by the Health Research Ethics Committee (CERES) of Paris Descartes University. They were paid 10€/hr. A trained psychologist conducted interviews based on the Mini International Neuropsychiatric Interview (Lecrubier et al., 1997) in order to ascertain that no participants met criteria for pathological anxiety disorders. None of them had a history of psychiatric disorders, neurological disorders or was currently undergoing medical treatment.

Table 1

Participants' Characteristics

Variable	All participants	NF group	CF group
Number of individuals	$N = 22$	$n_{NF} = 10$	$n_{CF} = 12$
Number of female [a]	14	5	9
Age ($M \pm SD$)[a]	22.95 ± 2.57	22.50 ± 2.12	23.33 ± 2.93
Crowd fear score ($M \pm SD$)[b]	16.64 ± 11.92	4.20 ± 1.81	27.00 ± 3.28
Range	[0.00; 33.00]	[0.00 ; 6.00]	[23.00 ; 33.00]
Trait anxiety score ($M \pm SD$)[c]	45.73 ± 8.60	41.30 ± 8.94	49.42 ± 6.56
Liebowitz social anxiety scale			
Anxiety sub-score ($M \pm SD$)[d]	25.45 ± 12.38	16.60 ± 6.29	32.83 ± 11.38
Avoidance sub-score ($M \pm SD$)[e]	20.02 ± 10.94	13.25 ± 5.36	25.67 ± 11.34

[a]Both groups were similar in terms of ratio of female (χ^2 test with Yates correction: $\chi^2(1) = 0.59$, $p = .442$) and age (the variable age deviated from normality hence a non-parametric Mann-Whitney U test was conducted: $U = 54.50$, $p = .380$).
[b]The crowd fear scores significantly deviated from a normal distribution within each group, for which reason a non-parametric test was conducted. The crowd fear score was significantly different between groups (Mann-Whitney U test: $U = 0.00$, $p < .001$).
[c]The trait anxiety scores assessed with the trait portion of the State Trait Anxiety Inventory (STAI) (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) was significantly different between groups (T test: $t(20) = 2.45$, $p = .023$, $d = 1.05$).
[d]The anxiety sub–score of the Liebowitz social anxiety scale (Liebowitz, 1987) was significantly different between groups (T test: $t(20) = 4.02$, $p < .001$, $d = 1.72$).
[e]The variance of the avoidance sub-scores was different between groups ($F(11, 9) = 4.66$, $p = .028$) hence a non-parametric test was conducted. The avoidance sub-score of the Liebowitz social anxiety scale was significantly different between groups (Mann-Whitney U test: $U = 24.50$, $p = .021$).

2.2. Virtual reality setup

The virtual reality setup was installed in an acoustically damped and soundproof recording studio. Participants wore polarized stereoscopic viewing glasses, headphones and a tracking device and were equipped with a 3D mouse (see Figure 1).

The visual scenes were presented on a 300 x 225-cm^2 stereoscopic passive screen, corresponding to 81.85 x 66.07 degrees at the viewing distance of 1.73 m, and were projected with two F2 SXGA + Projection Design projectors. The auditory scenes were presented through Sennheiser HD650 headphones. The sound stimuli were processed in real-time with the spatialization library Spat~ (Carpentier, Noisternig, & Warusfel, 2015) and rendered in binaural mode using a set of non-individual Head Related Transfer Functions (HRTF) selected from the LISTEN HRTF database (http://recherche.ircam.fr/equipes/salles/listen/). In binaural mode over headphones, the 3D location of a virtual sound source can be manipulated by rendering auditory cues such as inter-aural level and time differences as well as frequency filtering. The set of HRTF had been previously selected as best-fitting HRTF for a majority of participants in different experiments involving binaural rendering (see Moeck et al., 2007; Sarlat, Warusfel, & Viaud-Delmon, 2006). The 3D mouse was used by participants to navigate in the virtual environment. Moreover, head movements of the user were tracked using an OptiTrack optical system so

that visual stereo and 3D sounds were appropriately rendered with respect to the users' orientation. The voices of the humanoids were processed according to their distance and direction with regard to the current location and orientation of the user in the virtual environment. An ambient audio environment, composed of far field sound events (birds, bell, etc...) and recorded in ambisonic format, was rendered according to the current orientation of the user.

Fig. 1 Virtual reality setup.
A participant equipped with polarized glasses, headphones, a tracking device and a 3D mouse (left). A participant immersed in the virtual environment (right).

2.3.   Virtual environment and Crowd stimuli

The virtual environment reproduces the outdoor environment of the Trinity College Dublin campus composed of buildings, alleys and vegetation. Animated virtual individuals, referred to as humanoids, can be placed in this environment. The auditory component of the virtual environment consists of human speech and of an ambient audio environment composed of bird sounds and of urban activity.

In the virtual scene, in which participants navigated, twelve crowds of 96 animated humanoids (see example in Figure 2) were distributed in the virtual environment. They were each at a fixed location in the environment and could be presented through the visual channel only (unimodal visual –V–: a silent crowd), through the auditory channel only (unimodal auditory –A–: a crowd with talking humanoids that is hidden by a large canvas) or through both the auditory and visual channels (bimodal auditory-visual –AV–: a crowd with talking humanoids). The crowds were composed of an equal number of males and females humanoids, which were equally distributed in the right and left hemi-space of the participants' field of view as the latter approached. Twenty-six of the humanoids were talking in the auditory and auditory-visual crowds with an identical amount of female talkers and male talkers. Furthermore, no humanoid was oriented toward the user and humanoids

expressed neutral emotion. This was done to prevent from eliciting affective reactions unrelated to the perception of the crowd per se with the perception of direct gaze (e.g. Conty et al., 2010; Hietanen, Leppänen, Peltola, Linna-aho, & Ruuhiala, 2008) or with the perception of a panicked or threatening group of people (Huis in 't Veld & De Gelder, 2015). The use of a group of 96 humanoids was determined in a preliminary test in virtual reality, in which twelve participants judged the numerosity of groups of humanoids differing in size and presented through the auditory, the visual and both the auditory and visual channels (Taffou, Ondrej, O'Sullivan, Warusfel, & Viaud-Delmon, in preparation). The group of 96 humanoids was similarly considered as a crowd in each of the three sensory conditions.

Fig. 2 A crowd stimulus of 96 humanoids at 8m (left) and 2m (right) distance

Little numbered yellow flags were used as beacons to guide participants along the path to be explored. The four copies of the different crowd stimuli (A, V, AV) were presented in a pseudo-randomized manner along the path to be explored as described in Table 2. Little numbered red flags were used to help participants stop at 8m and at 2m from the crowd stimuli. These distances were chosen as they are considered to be respectively outside the public space, i.e. far, and within the social space, i.e. close, as described by Hall's proxemics framework (Hall, 1963, 1966). Moreover, given that near space is the area that individual humans actively maintain around themselves, into which other humans can not intrude without arousing discomfort (Hayduk, 1978), we had verified that CF and NF participants reported at least a slight discomfort at 2m from the crowds as an evidence for their location in near (peri-personal) space. The intensity of the different sound events was rendered according to natural attenuation law over distance (6dB drop per doubling of distance), i.e. the sound of a speaking humanoid at 2m from the participant would sound 12dB louder than at 8m. Concerning the visual component of the crowd stimuli, Fig. 2 exemplifies the different views at the 8- and 2-m flags. In order to facilitate the precision of participants' pause at the red flags, their texture changed. This texture change was triggered by the participants' proximity to the target position. Several additional humanoids, which were not part of the crowd stimuli, were also distributed in the virtual environment. They were added to fill the scene and increase realism but were not perceived by participants when the latter stopped at 8m and 2m from the crowd stimuli.

Table 2

Order of stimuli presentation during navigation

Order	Crowd stimuli	Repetition	Abbreviation
1	Auditory-visual	1	AV-1
2	Visual	1	V-1
3	Auditory	1	A-1
4	Auditory	2	A-2
5	Visual	2	V-2
6	Visual	3	V-3
7	Auditory-visual	2	AV-2
8	Auditory-visual	3	AV-3
9	Visual	4	V-4
10	Auditory	3	A-3
11	Auditory-visual	4	AV-4
12	Auditory	4	A-4

2.4. Procedure

Participants first completed the state portion of the State Trait Anxiety Inventory (Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983) before navigating in the auditory-visual virtual environment. Then, in order to become acquainted with the equipment and the navigation mode, the participant went through a training immersion in a humanoid-free version of the virtual scene. During this immersion, participants were also trained to follow a path to be explored by using numbered flags as guides as well as to stop at the flags if they were red (but not if they were yellow) and wait for the experimenter's instructions. The experimenter interacted with the participant in order to assist him/her in his/her first navigation.

After the training, each participant started the exploration of the virtual scene, which sought to measure participant's negative emotional experience at different distances from the auditory and visual crowd stimuli. The exploration began at the entryway of the virtual campus. The participant was instructed to follow numbered flags in order to explore the virtual scene. It was explained to him/her that, as in the training immersion, two kinds of flags could be found in the scene: yellow flags, which only serve to guide them along the path and red flags, at which they had to stop and wait for the experimenter's instructions. Each participant was informed that he/she would encounter several crowds along the exploration track and that some of the red flags could be placed quite close to the crowds. Participants encountered the different crowd stimuli in the order described before. For each stimulus, two red flags were used to place participants at 8m and 2m from the crowd. When encountering a crowd stimulus, participants had to first rate their discomfort level at 8m from the crowd (FAR condition) and then at 2m from the crowd (CLOSE condition). Discomfort ratings were collected using Subjective Unit of Distress (SUD; (Wolpe, 1973)), which is a standard self-report measurement of discomfort level on a 0-100 point scale. Four SUDs were collected in response to each different crowd stimuli (A, V, AV) and at each distance condition (CLOSE/FAR) for a total of 24 SUD measures.

After the navigation in the virtual scene, participants completed the presence questionnaire from the I-group (Schubert, Friedmann, & Regenbrecht, 2001), which evaluated their experience of presence in the virtual environment, and a 22-item cybersickness scale (Viaud-Delmon, Ivanenko, Berthoz, & Jouvent, 2000), which

assessed their level of discomfort linked to the use of the virtual reality setup. Then, they completed a second state portion of the STAI.

3. Results

One individual from the CF group did not complete the protocol because of manifestations of the autonomic nervous system related to VR (cybersickness). She stopped during training. Her score on the cybersickness scale was 32. All NF individuals completed the protocol. The following analyses were conducted on the remaining 21 participants (10 NF and 11 CF).

3.1. Questionnaire measures

We conducted an ANOVA with the between subject factor of GROUP (NF, CF) and the within subject factor of TIME (before, after) on the participants' state anxiety scores (see Table 3) to investigate the possible effect of our experimental protocol on participants state anxiety. The main effect of GROUP was significant, $F(1,19) = 8.25$, $p = .010$, $\eta_p^2 = .302$: state anxiety scores were higher in the CF group than in the NF group. There was no effect of neither the factor TIME, $F(1,19) = 0.112$, $p = .741$, $\eta_p^2 = .006$, nor the interaction GROUP*TIME, $F(1,19) = 0.148$, $p = .705$, $\eta_p^2 = .008$, on state anxiety scores suggesting that our experimental protocol had no effect on participants state anxiety.

Table 3

Individual Questionnaire Measures

ID	State anxiety 1	State anxiety 2	Cybersickness	Presence
Possible range	[20 –80]		[0 –88]	[0 –84]
Non Fearful group				
NF-1	35	35	0	48
NF-2	29	27	1	38
NF-3	24	23	0	48
NF-4	27	24	2	65
NF-5	26	25	0	41
NF-6	29	31	1	36
NF-7	24	20	0	66
NF-8	23	23	1	32
NF-9	34	40	13	14
NF-10	27	31	10	44
M ± SD	27.80 ± 4.08	27.90 ± 6.24	2.80 ± 4.69	43.20 ± 15.29
Crowd Fearful group				
CF-1	45	47	4	64
CF-2	32	37	4	36
CF-3	26	28	7	57
CF-4	27	35	5	48
CF-5	51	21	20	65
CF-6	34	50	2	53
CF-7	38	32	3	41
CF-8	38	44	5	41
CF-9	28	28	17	67
CF-10	31	27	15	20
CF-11	42	27	1	61
M ± SD	35.64 ± 7.99	34.18 ± 9.37	7.55 ± 6.58	50.27 ± 14.66

We conducted two-tailed non-parametric Mann–Whitney U tests to compare cybersickness and presence scores between groups because the distribution of these scores significantly deviated from a normal distribution. CF group's scores on the cybersickness questionnaire were higher than NF group's scores, U= 19.00, p = .012. This result is coherent with previous data showing that anxious or fearful participants experienced more severe cybersickness (e.g. Taffou et al., 2013; Viaud-Delmon, Warusfel, Seguelas, Rio, & Jouvent, 2006). There was no difference in presence scores between the two groups, U= 40.00, p = .304. The presence scores revealed that participants globally had the feeling of being in the campus during the experimental navigation immersion. However, two participants (one NF and one CF) had very low scores. The negative emotional experience they reported with SUDs in response to the crowd stimuli encountered during navigation was quite low, for both the NF participant (M = 0.00, SD = 0.00) and even for the CF participant (M = 0.83, SD = 2.82). This is coherent with previous data showing a link between presence and emotion in VR (Bouchard, St-Jacques, Robillard, & Renaud, 2008; Ling, Nefs, Morina, Heynderickx, & Brinkman, 2014; Riva et al., 2007; Robillard, Bouchard, Fournier, & Renaud, 2003).

3.2.   Navigation

Each participant managed to come as close as 2m from each stimulus of the virtual scene. As expected with the groups' constitution and definitions, crowds induced negative emotional experience in the CF group (M = 35.22, SD = 26.36) whereas they were not aversive stimuli in the NF group, who barely experienced very slight discomfort and reported very low SUDs (M = 4.16, SD = 4.54). Thus, the statistical analyses were done separately for each group. Within both groups (NF/CF), mean SUDs reported in response to unimodal stimuli (A-1, A-2, A-3, A-4, V-1, V-2, V-3, V-4) and to bimodal stimuli (AV-1, AV-2, AV-3, AV-4) were calculated for both DISTANCE conditions (close/far). The SUDs reported by the NF and by the CF groups in response to A, V and AV stimuli were not different between the four repetitions of each stimulus type (one way ANOVAs, p > .113 in all cases). Then, within each group, we conducted an ANOVA on the mean SUDs with the within subject factors SENSORY MODALITY (unimodal/bimodal) and DISTANCE (close/far).

In the CF group, the main effect of SENSORY MODALITY was significant, $F(1, 11) = 14.52$, p = .003, η_p^2 = .592, as well as the main effect of DISTANCE, $F(1, 10) = 22.01$, p < .001, η_p^2 = .688. The two-way interaction SENSORY MODALITY* DISTANCE was also significant, $F(1, 10) = 10.92$, p = .008, η_p^2 = .522, suggesting that, in the CF group, SUDs were differently modulated by whether crowds were unimodal or bimodal depending on the position of the crowds in close or far space. As the left panel of Figure 3 shows, CF participants reported significantly higher SUDs in response to bimodal compared to unimodal crowd stimuli when they were at a close distance from the crowds (post hoc Tukey's HSD test: p < .001). Contrastingly, there was no significant difference between SUDs reported in response to unimodal and bimodal crowd stimuli when CF participants were at a far distance from them (post hoc Tukey's HSD test: p = .097).

In the NF group (see right panel of Figure 3), the main effect of DISTANCE was significant, $F(1, 9) = 8.65$, p = .016, η_p^2 = .490, suggesting that SUDs reported by NF participants was influenced by the position of the crowds in close or far space. SUDs in response to close crowd stimuli were higher than SUDs in response to far crowd stimuli in the NF group. There was no significant effect of SENSORY MODALITY, $F(1, 9) = 1.75$, p = .218, η_p^2 = .163, or of the two-way interaction SENSORY MODALITY* DISTANCE, $F(1, 9) = 2.53$, p = .146, η_p^2 = .219, in the NF group.
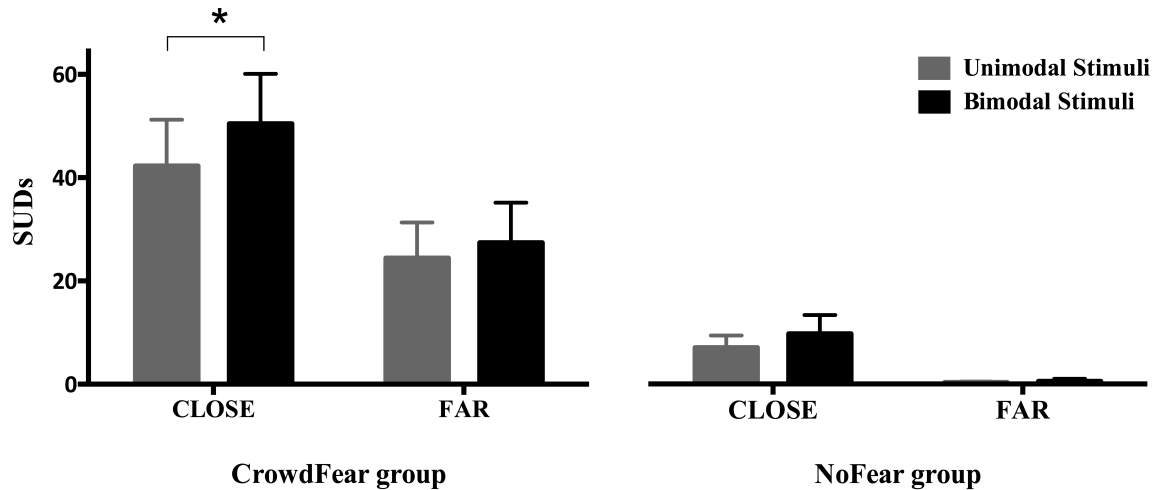
Fig. 3 Effect of bimodal crowd stimuli on negative emotional experience according to their spatial distance.
Mean SUDs (± SE) reported by the crowd-fearful (CF) group and the non-fearful (NF) group in response to unimodal (auditory or visual; black bars) and bimodal (auditory-visual; grey bars) crowd stimuli at 2m (CLOSE) and 8m (FAR) distances during the navigation. The bimodal crowd stimuli amplified the CF group's negative emotional experience when located at a close distance from participants.

Hence, bimodal crowds amplify negative emotional experience when they are considered as aversive (in the CF group) and when they are located at a close distance (2m) from the perceiver. To further investigate this effect of sensory modality in the close condition, we compared the mean SUDs in response to the auditory crowds on one hand and the mean SUDs in response to the visual crowds on the other hand to the mean SUDs in response to auditory-visual crowds. As illustrated on the left panel of Figure 4, CF participants reported significantly lower SUDs in response to visual crowd stimuli compared to auditory-visual crowd stimuli, $t(10) = 4.14$, $p = .002$, $d = 1.25$, when they were at a close distance from the crowds. There was no difference in SUDs reported by the CF group between auditory and auditory-visual close crowd stimuli, $t(10) = 1.79$, $p = .104$, $d = 0.54$. However, there was no significant difference between SUDs in response to auditory close crowd stimuli and visual close crowd stimuli ($t(10) = 1.54$, $p = .154$, $d = 0.46$). In the NF group, as the right panel of Figure 4 shows, there was no significant difference between SUDs reported in response to auditory and auditory-visual close crowd stimuli nor between SUDs reported in response to visual and auditory-visual close crowd stimuli ($p > .068$ in both cases). Hence, bimodal auditory-visual presentation of closed feared stimuli amplifies negative emotional experience compared with the corresponding unimodal presentation and auditory stimulation seems to have a specific role in this multisensory effect.
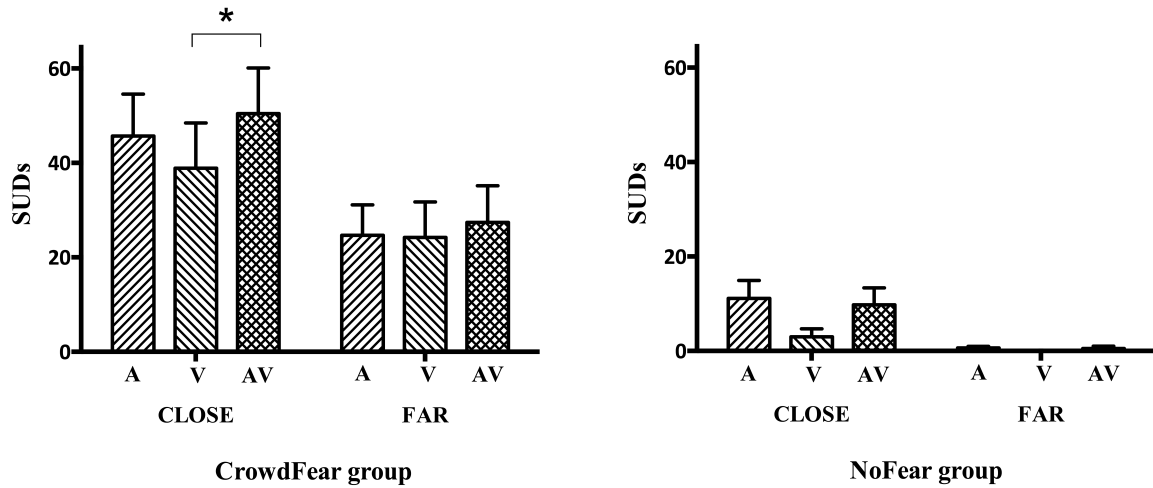
Fig. 4 Negative emotional experience in response to A, V and AV crowd stimuli according to their spatial distance.
Mean SUDs (± SE) reported by the crowd-fearful (CF) group and the non-fearful (NF) group in response to auditory (blue bars), visual (red bars) and auditory-visual (purple bars) crowd stimuli at 2m (CLOSE) and 8m (FAR) distances during the navigation. At close distances, CF group's negative emotional experience in response to visual crowd stimuli was lower than in response to auditory-visual crowd stimuli.

4. Discussion

The goal of this study was to investigate how multisensory presentation of aversive stimuli influences the negative emotional experience induced in the perceiver and to test whether the distance between the stimulus and the perceiver modulates this influence. Our findings show that when a stimulus is perceived as neutral, there is no modification of the emotional experience linked to the sensory presentation. However, when a stimulus is perceived as aversive, as it is the case in the CF group, the emotional experience is enhanced by auditory-visual presentation at a close, but not at a far distance from the perceiver. Our findings also suggest a specific role for auditory information in this multisensory effect.

Crowd-fearful (CF) participants reported more intense feelings of discomfort in response to bimodal crowds as compared with unimodal crowds when standing at a close distance from them. This finding suggests that the multiplicity of sensory cues conveyed by emotional stimuli modulates affective processing stage of subjective experience of emotion but only if the emotional stimuli is at a close distance. This is consistent with previous results demonstrating that affective events conveying multisensory emotional cues increase the emotional experience induced in the perceiver (Baumgartner et al., 2006; Christensen et al., 2014; Taffou et al., 2013; Vines et al., 2011, 2006) and refines them by showing that the emotional effect of multisensory integration is influenced by the distance between the perceiver and the affective event.

The effect of multisensory presentation on emotional experience that we observed with close stimuli, i.e. a difference between the SUDs in the bimodal (AV) condition and the SUDs in the corresponding unimodal (V and A) condition, could be driven by the stronger responses induced by the auditory stimuli or could be due to a combination of auditory and visual information. Indeed, whereas unimodal V close crowds induced lower emotional experience than bimodal AV close crowds, it was not the case with unimodal A close crowds. This

effect could be linked to the higher restriction of the perception of people via vision than via audition at close distances to the crowds. It could also be linked to the fact that the production of a conscious emotional experience involves emotional regulation processes, which take context into account. In the A condition, even though participants received only auditory information from the crowds, they were aware that a visual component was implicitly linked to the crowd but that, due to the big canvas blocking their vision, they could not perceive it. Furthermore, we cannot know whether the canvas it self had an effect on the emotional response. In the V condition, participants did not perceive auditory information from the crowd: it was clear that it was a silent crowd. In this sense, the V condition might be seen as more unimodal than the A condition, which is implicitly bimodal. Therefore, although our study show that auditory-visual emotional stimuli amplify negative emotional experience, the specific role of auditory information in this effect has to be further investigated. Moreover, whether this modulation of the subjective experience of emotion is linked to the interaction of multisensory processes with only the stage of subjective experience of emotion, or also with the first stages of affective processing remains to be understood. In order to address these questions, future research should focus on designing middle-ground paradigms involving naturalistic stimulation embedded in an ecological context while conceding on the interactivity of stimuli in order to enable the monitoring of the time-course of multisensory affective processing from early stages on with electrophysiological techniques such as electroencephalography as well as the measure of emotional responses such as heart rate or galvanic skin resistance changes along with the collect of subjective experience of emotion.

At a farther distance, auditory-visual crowds did not increase discomfort in CF participants. At an 8m-distance, although the crowds still induced negative feelings in CF participants, the intensity of the negative feelings was not influenced by the multisensory presentation of the stimuli. Aversive stimuli would evoke an emotional experience in far space, which would however not be potentiated by the multiplicity of sensory cues. This finding that multisensory integration has a stronger effect for emotional cues present at close (2m) versus far distances (8m) suggests that multisensory integration processes could be specific to near space. At close distances, the quantity of sensory information received from external events is more important and they can be differently appraised in comparison to far distances. For example, with crowd stimuli, the numerosity of people may be evaluated as higher at close compared to far distances leading to higher crowding feelings (Hayduk, 1983). Moreover, the combination of different sensory emotional cues conveyed by an event contributes to improve emotion perception in terms of accuracy and rapidity (Collignon et al., 2008; De Gelder et al., 1999; De Gelder & Vroomen, 2000; Dolan et al., 2001; Föcker et al., 2011; Kreifelts et al., 2007; Li et al., 2013; Pourtois et al., 2005; Tanaka et al., 2010; Van den Stock et al., 2008). Thus, given that close events represent more of a potential threat than distant ones, this is congruent with a higher impact of multisensory integration processes on affective processing with emotional cues conveyed by affective events located close to the body. Many studies have brought evidence for a stronger effect of multisensory integration of neutral cues with stimuli located near the body. They mostly investigated visuo-tactile (e.g. Holmes, Sanabria, Calvert, & Spence, 2007; Spence, Pavani, & Driver, 2004) and auditory-tactile (e.g. Kitagawa, Zampini, & Spence, 2005; Tajadura-Jiménez et al., 2009; Zampini, Torresan, Spence, & Murray, 2007) interactions and examined sensory spatial cues' incongruence or redundancy effects on task performance as an indicator for multisensory integration of neutral information. They all revealed evidence of a stronger effect of auditory-tactile or visuo-tactile integration of neutral cues when both cues were located close to the body, as compared to farther distances. In contrast to

tactile stimuli, auditory and visual stimuli can both be positioned in near and far spaces. In a recent study, Van der Stoep et al. observed a strong effect of auditory-visual integration on speed of detection with neutral cues located at 2m and at 0.8m (Van der Stoep, Van der Stigchel, Nijboer, & Van der Smagt, 2015). Our findings are coherent with these previous results obtained with non-emotional stimuli and suggest that there is also a specific multisensory integration for emotional cues in near space. Future studies should investigate the neural correlates of the interaction between affective, multisensory and spatial representation processes in order to test whether the convergence occurs within subcortical structures such as the superior colliculus, the amygdala or the thalamus or whether it involves cortical structures such as the parietal or the insular cortex, or both.

The NF group only experienced really slight negative emotion in response to crowds located at 2m. Their emotional experience with close crowds was less intense than the one of CF participants with far crowds, whatever the sensory presentation. This suggests that spatial closeness and multisensory presentation of stimuli is a non-sufficient condition for the emergence of negative emotional experience in the perceiver. We think that in addition to these extrinsic qualities, a negative emotional value has to have been conferred to the stimuli. Here, the intrinsic factor that is the sensitivity to the fear of crowds was used to manipulate the attribution of an emotional valence to the crowd stimuli. Crowds are not aversive for NF participants. Consequently, it is not an aversive stimulation, which enters near space. When not aversive, some stimuli – attractive stimuli – can be non-invasive and even desirable in near space. It is especially true when the stimulus is a human because the positive feelings evoked by intimacy with romantic partners or family members often originate from social interaction in near space (Aiello, 1987; Hall, 1966). Consequently, the fact that the spatial and sensory presentation of the crowds did not have the same effect in the NF and the CF group is coherent with near space being an area allowing for the preparation and coordination of defensive behaviors (Graziano & Cooke, 2006). Moreover, it is also consistent with the fact that multisensory integration is thought to stimulate detection and to allow responding behaviorally to a stimulus that we would have missed if sensed through only one sensory channel (Bolognini, Frassinetti, Serino, & Làdavas, 2005; Lovelace, Stein, & Wallace, 2003; Serino, Pizzoferrato, & Làdavas, 2008; Stein & Stanford, 2008). Thus, the fact that the extrinsic factors of multisensory presentation and spatial closeness are not sufficient to give rise to a subjective experience of negative emotion in response to crowds of humans may be linked to a social purpose. In the present study, we first focused on two factors that modulate the emotional impact of crowds i.e. distance and sensory presentation but we think that other characteristics such as gaze, facial and body expression of the people constituting the crowd and also the dynamic of the crowd (people standing or walking) can influence the negative emotional experience that the crowd can induce. Although our study did not allow investigating all these characteristics, it is definitely something that should be addressed in future research.

Altogether, our findings suggest that spatial closeness allows multisensory processes to modulate the intensity of the emotional experience induced by aversive stimuli. Nevertheless, the specific role of auditory stimulation must be investigated in order to better understand this effect. We think that the close links between multisensory, affective and space representation processes, which we further expose in this study, serve a social purpose. The required implementation of defensive behaviors in response to aversive events positioned to threaten individual's comfort may be achieved through an emotional effect of multisensory integration of cues located within peri-personal space.

5. References

Aiello, J. R. (1987). Human spatial behavior. In D. Stokols & I. Altman (Eds.), *Handbook of environmental psychology* (pp. 389–504). New York, NY: John Wiley & sons.

Baumgartner, T., Lutz, K., Schmidt, C. F., & Jäncke, L. (2006). The emotional power of music: how music enhances the feeling of affective pictures. *Brain Research*, *1075*(1), 151–64. http://doi.org/10.1016/j.brainres.2005.12.065

Bolognini, N., Frassinetti, F., Serino, A., & Làdavas, E. (2005). "Acoustical vision" of below threshold stimuli: interaction among spatially converging audiovisual inputs. *Experimental Brain Research*, *160*(3), 273–82. http://doi.org/10.1007/s00221-004-2005-z

Bouchard, S., St-Jacques, J., Robillard, G., & Renaud, P. (2008). Anxiety Increases the Feeling of Presence in Virtual Reality. *Presence: Teleoperators and Virtual Environments*.

Brozzoli, C., Makin, T. R., Cardinali, L., Holmes, N. P., & Farnè, A. (2012). Peripersonal space. In M. M. Murray & M. T. Wallace (Eds.), *The Neural Bases of Multisensory Processes*. Baco Raton, FL: CRC Press.

Carpentier, T., Noisternig, M., & Warusfel, O. (2015). Twenty years of Ircam Spat: looking back, looking forward. In *International Computer Music Conference Proceedings*.

Christensen, J. F., Gaigg, S. B., Gomila, A., Oke, P., & Calvo-Merino, B. (2014). Enhancing emotional experiences to dance through music: the role of valence and arousal in the cross-modal bias. *Frontiers in Human Neuroscience*, *8*(October), 1–9. http://doi.org/10.3389/fnhum.2014.00757

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M., & Lepore, F. (2008). Audio-visual integration of emotion expression. *Brain Research*, *1242*, 126–35. http://doi.org/10.1016/j.brainres.2008.04.023

Conty, L., Russo, M., Loehr, V., Hugueville, L., Barbu, S., Huguet, P., … George, N. (2010). The mere perception of eye contact increases arousal during a word-spelling task. *Social Neuroscience*, *5*(2), 171–186.

Cowey, A., Small, M., & Ellis, S. (1994). Left visuo-spatial neglect can be worse in far than in near space. *Neuropsychologia*, *32*(9), 1059–1066.

Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous system. *Brain Research. Brain Research Reviews*, *26*(2-3), 83–6.

De Gelder, B., Böcker, K. B., Tuomainen, J., Hensen, M., & Vroomen, J. (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. *Neuroscience Letters*, *260*(2), 133–6.

De Gelder, B., & Vroomen, J. (2000). The perception of emotion by ear and by eye. *Cognition and Emotion*, *14*(3), 289–311.

Diederich, A., & Colonius, H. (1987). Intersensory facilitation in the motor component? *Psychological Research*, *49*, 23–29.

Dolan, R. J., Morris, J. S., & De Gelder, B. (2001). Crossmodal binding of fear in voice and face. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(17), 10006–10. http://doi.org/10.1073/pnas.171288598

Dosey, M. A., & Meisels, M. (1969). Personal space and self-protection. *Journal of Personality and Social Psychology*, *11*(2), 93–7.

Ferri, F., Tajadura-Jiménez, A., Väljamäe, A., Vastano, R., & Costantini, M. (2015). Emotion-inducing approaching sounds shape the boundaries of multisensory peripersonal space. *Neuropsychologia*, *70*(April), 468–75. http://doi.org/10.1016/j.neuropsychologia.2015.03.001

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

Giray, M., & Ulrich, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology. Human Perception and Performance*, *19*(6), 1278–1291.

Gondan, M., Lange, K., Rösler, F., & Röder, B. (2004). The redundant target effect is affected by modality switch costs. *Psychonomic Bulletin & Review*, *11*(2), 307–13.

Gondan, M., Niederhaus, B., Rösler, F., & Röder, B. (2005). Multisensory processing in the redundant-target effect: a behavioral and event-related potential study. *Perception & Psychophysics*, *67*(4), 713–26.

Graziano, M. S. A., & Cooke, D. F. (2006). Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia*, *44*(6), 845–59. http://doi.org/10.1016/j.neuropsychologia.2005.09.009

Hagan, C. C., Woods, W., Johnson, S., Green, G. G. R., & Young, A. W. (2013). Involvement of Right STS in Audio-Visual Integration for Affective Speech Demonstrated Using MEG. *PloS One*, *8*(8). http://doi.org/10.1371/journal.pone.0070648

Hall, E. T. (1963). A system for the notation of proxemic behavior. *American Anthropologist*, *65*(5, Selected Papers in Method and Technique), 1003–1026.

Hall, E. T. (1966). *The Hidden Dimension*. New York, NY: Doubleday.

Halligan, P. W., & Marshall, J. C. (1991). Left neglect for near but not far space in man. *Nature*, *350*(6318), 498–500.

Hayduk, L. A. (1978). Personal space: An evaluative and orienting overview. *Psychological Bulletin*, *85*(1), 117–134. http://doi.org/10.1037//0033-2909.85.1.117

Hayduk, L. A. (1983). Personal space: Where we now stand. *Psychological Bulletin*, *94*(2), 293–335. http://doi.org/10.1037//0033-2909.94.2.293

Hietanen, J. K., Leppänen, J. M., Peltola, M. J., Linna-aho, K., & Ruuhiala, H. J. (2008). Seeing direct and averted gaze activates the approach-avoidance motivational brain systems. *Neuropsychologia*, *46*(9), 2423–2430.

Holmes, N. P., Sanabria, D., Calvert, G. A., & Spence, C. (2007). Tool-use: capturing multisensory spatial attention or extending multisensory peripersonal space? *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *43*(3), 469–89.

Huis in 't Veld, E. M. J., & De Gelder, B. (2015). From personal fear to mass panic: The neurological basis of crowd perception. *Human Brain Mapping*. http://doi.org/10.1002/hbm.22774

Kitagawa, N., Zampini, M., & Spence, C. (2005). Audiotactile interactions in near and far space. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *166*(3-4), 528–37. http://doi.org/10.1007/s00221-005-2393-8

Klasen, M., Chen, Y.-H., & Mathiak, K. (2012). Multisensory emotions: perception, combination and underlying neural processes. *Reviews in the Neurosciences*, *23*(4), 381–92. http://doi.org/10.1515/revneuro-2012-0040

Kokinous, J., Kotz, S. A., Tavano, A., & Schröger, E. (2015). The role of emotion in dynamic audiovisual integration of faces and voices. *Social Cognitive and Affective Neuroscience*, *10*(5), 713–20. http://doi.org/10.1093/scan/nsu105

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M., & Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: an event-related fMRI study. *NeuroImage*, *37*(4), 1445–56. http://doi.org/10.1016/j.neuroimage.2007.06.020

Laurienti, P. J., Kraft, R. A., Maldjian, J. A., Burdette, J. H., & Wallace, M. T. (2004). Semantic congruence is a critical factor in multisensory behavioral performance. *Experimental Brain Research*, *158*, 405–14. http://doi.org/10.1007/s00221-004-1913-2

Lecrubier, Y., Sheehan, D. V., Weiller, E., Amorim, P., Bonora, I., Sheehan, K. H., … Dunbar, G. C. (1997). The Mini International Neuropsychiatric Interview (MINI). A short diagnostic structured interview: Reliability and validity according to the CIDI. *European Psychiatry*, *12*(5), 224–231.

Li, Y., Long, J., Huang, B., Yu, T., Wu, W., Liu, Y., … Sun, P. (2013). Crossmodal Integration Enhances Neural Representation of Task-Relevant Features in Audiovisual Face Perception. *Cerebral Cortex*. http://doi.org/10.1093/cercor/bht228

Ling, Y., Nefs, H. T., Morina, N., Heynderickx, I., & Brinkman, W. P. (2014). A meta-analysis on the relationship between self-reported presence and anxiety in virtual reality exposure therapy for anxiety disorders. *PLoS ONE*, *9*(5).

Lourenco, S. F., Longo, M. R., & Pathman, T. (2011). Near space and its relation to claustrophobic fear. *Cognition*, *119*(3), 448–53. http://doi.org/10.1016/j.cognition.2011.02.009

Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, *17*(2), 447–53.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*(2), 247–79.

Miller, J. (1991). Channel interaction and the redundant-targets effect in bimodal divided attention. *Journal of Experimental Psychology. Human Perception and Performance*, *17*(1), 160–9.

Mobbs, D., Petrovic, P., Marchant, J. L., Hassabis, D., Weiskopf, N., Seymour, B., … Frith, C. D. (2007). When fear is near: threat imminence elicits prefrontal-periaqueductal gray shifts in humans. *Science (New York, N.Y.)*, *317*(5841), 1079–1083.

Moeck, T., Bonneel, N., Tsingos, N., Drettakis, G., Viaud-Delmon, I., & Alloza, D. (2007). Progressive Perceptual Audio Rendering of Complex Scenes. In *Proceedings of the 2007 symposium on Interactive 3D graphics and games, April 30-May 02, 2007*. Seattle, Washington.

Molholm, S., Ritter, W., Javitt, D. C., & Foxe, J. J. (2004). Multisensory Visual-Auditory Object Recognition in Humans: a High-density Electrical Mapping Study. *Cerebral Cortex*, *14*(4), 452–465. http://doi.org/10.1093/cercor/bhh007

Phillips, M. L., Drevets, W. C., Rauch, S. L., & Lane, R. (2003). Neurobiology of emotion perception I: the neural basis of normal emotion perception. *Biological Psychiatry*, *54*(5), 504–514. http://doi.org/10.1016/S0006-3223(03)00168-9

Pourtois, G., De Gelder, B., Bol, A., & Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain. *Cortex; a Journal Devoted to the Study of the Nervous System and Behavior*, *41*(1), 49–59.

Previc, F. H. (1998). The Neuropsychology of 3-D Space. *Psychological Bulletin*, *124*(2), 123–164.

Risberg, A., & Lubker, J. (1978). Prosody and speech-reading. *Quarterly Progress and Status Report Prosody and Speechreading*, *4*, 1–16.

Riva, G., Mantovani, F., Capideville, C. S., Preziosa, A., Morganti, F., Villani, D., … Alcañiz, M. (2007). Affective interactions using virtual reality: the link between presence and emotions. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *10*(1), 45–56.

Rizzolatti, G., Fadiga, L., Fogassi, L., & Gallese, V. (1997). The space around us. *Science (New York, N.Y.)*, *277*(5323), 190–1.

Robillard, G., Bouchard, S., Fournier, T., & Renaud, P. (2003). Anxiety and presence during VR immersion: a comparative study of the reactions of phobic and non-phobic participants in therapeutic virtual environments derived from computer games. *Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society*, *6*(5), 467–76. http://doi.org/10.1089/109493103769710497

Sambo, C. F., & Iannetti, G. D. (2013). Better safe than sorry? The safety margin surrounding the body is increased by anxiety. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *33*(35), 14225–30. http://doi.org/10.1523/JNEUROSCI.0706-13.2013

Sarlat, L., Warusfel, O., & Viaud-Delmon, I. (2006). Ventriloquism aftereffects occur in the rear hemisphere. *Neuroscience Letters*, *404*(3), 324–9. http://doi.org/10.1016/j.neulet.2006.06.007

Schubert, T., Friedmann, F., & Regenbrecht, H. (2001). The experience of presence: Factor analytic insights. *Presence Teleoperators and Virtual Environments*, *10*, 266–281.

Serino, A., Pizzoferrato, F., & Làdavas, E. (2008). Viewing a face (especially one's own face) being touched enhances tactile perception on the face. *Psychological Science*, *19*(5), 434–8. http://doi.org/10.1111/j.1467-9280.2008.02105.x

Spence, C., Pavani, F., & Driver, J. (2004). Spatial constraints on visual-tactile cross-modal distractor congruency effects. *Cognitive, Affective & Behavioral Neuroscience*, *4*(2), 148–169.

Spielberger, C. D., Gorsuch, R. L., Lushene, P. R., Vagg, P. R., & Jacobs, A. G. (1983). *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press.

Stein, B. E., & Stanford, T. R. (2008). Multisensory integration: current issues from the perspective of the single neuron. *Nature Reviews. Neuroscience*, *9*(4), 255–66. http://doi.org/10.1038/nrn2331

Suied, C., Bonneel, N., & Viaud-Delmon, I. (2009). Integration of auditory and visual information in the recognition of realistic objects. *Experimental Brain Research. Experimentelle Hirnforschung. Expérimentation Cérébrale*, *194*(1), 91–102. http://doi.org/10.1007/s00221-008-1672-6

Sumby, W. H., & Pollack, I. (1954). Visual Contribution to Speech Intelligibility in Noise. *The Journal of the Acoustical Society of America*, *26*(2), 212–215.

Taffou, M., Guerchouche, R., Drettakis, G., & Viaud-Delmon, I. (2013). Auditory–Visual Aversive Stimuli Modulate the Conscious Experience of Fear. *Multisensory Research*, *26*, 347–370. http://doi.org/10.1163/22134808-00002424

Taffou, M., Ondrej, J., O'Sullivan, C., Warusfel, O., & Viaud-Delmon, I. (submitted). Judging crowds' size by ear and by eye in virtual reality. *Journal on Multimodal User Interfaces*

Taffou, M., & Viaud-Delmon, I. (2014). Cynophobic Fear Adaptively Extends Peri-Personal Space. *Frontiers in Psychiatry*, *5*(September), 3–9. http://doi.org/10.3389/fpsyt.2014.00122

Tajadura-Jiménez, A., Kitagawa, N., Väljamäe, A., Zampini, M., Murray, M. M., & Spence, C. (2009). Auditory-somatosensory multisensory interactions are spatially modulated by stimulated body surface and acoustic spectra. *Neuropsychologia*, *47*(1), 195–203. http://doi.org/10.1016/j.neuropsychologia.2008.07.025

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E., & De Gelder, B. (2010). I feel your voice. Cultural differences in the multisensory perception of emotion. *Psychological Science*, *21*(9), 1259–62. http://doi.org/10.1177/0956797610380698

Vagnoni, E., Lourenco, S. F., & Longo, M. R. (2012). Threat modulates perception of looming visual stimuli. *Current Biology : CB*, *22*(19), R826–7. http://doi.org/10.1016/j.cub.2012.07.053

Van den Stock, J., Grèzes, J., & De Gelder, B. (2008). Human and animal sounds influence recognition of body language. *Brain Research*, *1242*, 185–90. http://doi.org/10.1016/j.brainres.2008.05.040

Van der Stoep, N., Van der Stigchel, S., Nijboer, T. C. W., & Van der Smagt, M. J. (2015). Audiovisual integration in near and far space: effects of changes in distance and stimulus effectiveness. *Experimental Brain Research*, (March 19), 1–14. http://doi.org/10.1007/s00221-015-4248-2

Viaud-Delmon, I., Ivanenko, Y. P., Berthoz, A., & Jouvent, R. (2000). Adaptation as a sensorial profile in trait anxiety: a study with virtual reality. *Journal of Anxiety Disorders*, *14*(6), 583–601.

Viaud-Delmon, I., Warusfel, O., Seguelas, A., Rio, E., & Jouvent, R. (2006). High sensitivity to multisensory conflicts in agoraphobia exhibited by virtual reality. *European Psychiatry : The Journal of the Association of European Psychiatrists*, *21*(7), 501–8. http://doi.org/10.1016/j.eurpsy.2004.10.004

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M., & Levitin, D. J. (2011). Music to my eyes: cross-modal interactions in the perception of emotions in musical performance. *Cognition*, *118*(2), 157–70. http://doi.org/10.1016/j.cognition.2010.11.010

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., & Levitin, D. J. (2006). Cross-modal interactions in the perception of musical performance. *Cognition*, *101*, 80–113. http://doi.org/10.1016/j.cognition.2005.09.003

Watson, R., Latinus, M., Noguchi, T., Garrod, O., Crabbe, F., & Belin, P. (2014). Crossmodal Adaptation in Right Posterior Superior Temporal Sulcus during Face-Voice Emotional Integration. *Journal of Neuroscience*, *34*(20), 6813–6821. http://doi.org/10.1523/JNEUROSCI.4478-13.2014

Wolpe, J. (1973). *The practice of behavior therapy (2nd ed.)*. New York (NY): Pergamon.

Zampini, M., Torresan, D., Spence, C., & Murray, M. M. (2007). Auditory-somatosensory multisensory interactions in front and rear space. *Neuropsychologia*, *45*(8), 1869–77. http://doi.org/10.1016/j.neuropsychologia.2006.12.004