BRILL
Multisensory Research 26 (2013) 347–370
brill.com/msr

Auditory–Visual Aversive Stimuli Modulate the Conscious Experience of Fear

Marine Taffou [1,2,*], **Rachid Guerchouche** [3], **George Drettakis** [3] and **Isabelle Viaud-Delmon** [1]

[1] UMR 9912 CNRS UPMC IRCAM, 1 place Igor Stravinsky, 75004 Paris, France
[2] Centre de Recherche de l'Institut du Cerveau et de la Moelle épinière (CRICM), UMR S975 UPMC, UMR 7225 CNRS, U975 INSERM, 47 Boulevard de l'Hôpital, 75013 Paris, France
[3] REVES INRIA, 2004 Route des Lucioles, 06902 Sophia Antipolis, France

Received 4 March 2013; accepted 17 June 2013

Abstract

In a natural environment, affective information is perceived *via* multiple senses, mostly audition and vision. However, the impact of multisensory information on affect remains relatively undiscovered. In this study, we investigated whether the auditory–visual presentation of aversive stimuli influences the experience of fear. We used the advantages of virtual reality to manipulate multisensory presentation and to display potentially fearful dog stimuli embedded in a natural context. We manipulated the affective reactions evoked by the dog stimuli by recruiting two groups of participants: dog-fearful and non-fearful participants. The sensitivity to dog fear was assessed psychometrically by a questionnaire and also at behavioral and subjective levels using a Behavioral Avoidance Test (BAT). Participants navigated in virtual environments, in which they encountered virtual dog stimuli presented through the auditory channel, the visual channel or both. They were asked to report their fear using Subjective Units of Distress. We compared the fear for unimodal (visual or auditory) and bimodal (auditory–visual) dog stimuli. Dog-fearful participants as well as non-fearful participants reported more fear in response to bimodal audiovisual compared to unimodal presentation of dog stimuli. These results suggest that fear is more intense when the affective information is processed *via* multiple sensory pathways, which might be due to a cross-modal potentiation. Our findings have implications for the field of virtual reality-based therapy of phobias. Therapies could be refined and improved by implicating and manipulating the multisensory presentation of the feared situations.

Keywords

Multisensory integration, emotion, fear, cynophobia, virtual reality, VRET

* To whom correspondence should be addressed. E-mail: marine.taffou@ircam.fr

DOI:10.1163/22134808-00002424

1. Introduction

Affective situations often deliver cues across multiple sensory modalities: when encountering an aggressive dog, the threat is perceived *via* both vision and audition. While affective processing has mostly been studied in one sensory modality at a time, an increasing number of studies have aimed at exploring how we deal with affective information coming from multiple senses. These studies mostly used affective faces paired with affective voices, since these stimuli represent a common and natural multisensory affective situation, in normal participants (Chen *et al.*, 2010; Collignon *et al.*, 2008; De Gelder *et al.*, 1999, 2002; De Gelder and Vroomen, 2000; Dolan *et al.*, 2001; Föcker *et al.*, 2011; Hagan *et al.*, 2009; Jessen and Kotz, 2011; Koizumi *et al.*, 2011; Kreifelts *et al.*, 2007; Massaro and Egan, 1996; Müller *et al.*, 2012; Pourtois *et al.*, 2000, 2005; Robins *et al.*, 2009; Tanaka *et al.*, 2010; Vroomen *et al.*, 2001) and patients with schizophrenia (De Gelder *et al.*, 2005; De Jong *et al.*, 2009, 2010), autism spectrum disorder (Magnée *et al.*, 2011), pervasive developmental disorders (Magnée *et al.*, 2007, 2008) or alcoholism (Maurage *et al.*, 2008). The combination of emotionally-congruent facial expression and prosody facilitates emotional judgment of negatively- and positively-valenced stimuli (Collignon *et al.*, 2008; Dolan *et al.*, 2001; Föcker *et al.*, 2011; Kreifelts *et al.*, 2007; Massaro and Egan, 1996) and seems to be a mandatory process, unconstrained by attentional resources (Collignon *et al.*, 2008; De Gelder and Vroomen, 2000; Föcker *et al.*, 2011; Vroomen *et al.*, 2001).

However, these studies have concentrated on the first steps of affective processing. The processing of an affective stimulus comprises several stages from the evaluation of the affective significance of the stimulus, to the conscious experience of emotion also called feeling, and the regulation of the emotional response (Damasio, 1998; Phillips *et al.*, 2003; Rudrauf *et al.*, 2009). If the first stages of affective processing have been shown to be influenced by multisensory information, their effects on the conscious experience of emotion remain to be elucidated.

Few studies have explored the influence of combined presentation of auditory and visual stimuli on feeling. Aesthetic experience has been shown to be enhanced in response to auditory–visual compared to unimodal presentation of musical performances (Vines *et al.*, 2006, 2011). An increased experience of emotion has also been found in response to positive and negative non-natural pairs of affective pictures and music, when compared to the response to affective pictures only (Baumgartner *et al.*, 2006). It is not yet clear whether the multisensory presentation of stimuli impacts the conscious experience of emotion.

In this study, our goal was to manipulate the presentation of auditory and visual aversive stimuli in order to investigate whether the multisensory presen-

Table 1.
Abbreviations

BAT	Behavioral Avoidance Test
nSCL	Normalized Skin Conductance Level
SCL	Skin Conductance Level
SUD	Subjective Unit of Distress
VE	Virtual Environment
VR	Virtual Reality

tation influences the conscious experience of fear. Since the auditory–visual presentation of affective stimuli facilitates affective judgments (Collignon *et al.*, 2008; Dolan *et al.*, 2001; Föcker *et al.*, 2011; Kreifelts *et al.*, 2007; Massaro and Egan, 1996), we hypothesized that it would also lead to an enhanced fear. To explore the effect of the multisensory presentation of aversive stimuli on fear, we used a fully immersive virtual reality setup system to display dog stimuli within auditory–visual virtual environments (VEs; abbreviations are listed in Table 1). Dogs are considered as fear-relevant stimuli for humans in general and can be genuinely aversive and fearful for a subset of individuals sensitive to the fear of dogs. Furthermore, this stimulus can convey affective information *via* both auditory and visual pathways. Virtual reality integrates real-time computer graphics, body tracking devices and visual and auditory displays to immerse a user in a computer-generated VE. The setting in which the user performs an action can be controlled by the experimenter, recorded and measured. The unique features and flexibility of VR give it extraordinary potential for use in multisensory integration research. Immersing a participant in a VE enables biologically-relevant auditory–visual stimuli to be presented embedded within a natural context as well as to manipulate the sensory characteristics of the stimuli (Bohil *et al.*, 2011).

A sample of healthy participants sensitive to the fear of dogs and a sample of healthy participants non-sensitive to the fear of dogs were exposed to virtual dog stimuli and reported their fear. We expected that the dog-fearful participants would report higher fear in response to bimodal (auditory–visual) compared to unimodal dog stimuli. For the non-fearful participants, dogs are fear-relevant but not fearful or aversive. Hence, we expected that, in contrast to the dog-fearful participants, they would not experience any feeling of fear in response to the dog stimuli.

We presented the supposedly less fearful (unimodal) stimuli before the supposedly most fearful (bimodal) stimuli to avoid implosive, long-lasting experience of fear and the subsequent saturation effect on feeling (e.g. Nesse *et al.*, 1980; Pitman *et al.*, 1996), which could mask the phenomenon of interest. We also distributed the dog stimuli within the VEs to prevent any overlap of fear (Garrett and Maddock, 2001). The participants' task was to explore these VEs in order to find an auditory–visual frog. Thus, we created a paradigm aiming at investigating the conscious experience of fear in the most appropriate and natural manner. We also measured the skin conductance level (SCL) as an indicator of participants' arousal state during the presentation of our fearful stimuli. This measure allowed us to explore whether bimodal as compared to unimodal stimuli would evoke stronger non-conscious fear. If this is the case, bimodal stimuli would further increase emotionally-induced defense engagement and thus further enhance autonomic responses such as the SCL (Bradley *et al.*, 2001; Kreibig, 2010).

2. Methods

The experiment was composed of two sessions, which took place on two different days. In the first session, participants were invited to take part in a twenty minute long diagnostic interview, based on the Mini International Neuropsychiatric Interview, with a clinical psychologist. This interview was conducted to make certain that no participant met criteria for pathological anxiety disorders. The second session consisted of several immersions in four different VEs and the completion of several questionnaires. The total duration of the second session was an hour and a half.

During the second session, the procedure was as follows: each participant was first submitted to a Behavioral Avoidance Test (BAT) in a VE (see Mühlberger *et al.*, 2008 for another example of a BAT conducted in virtual reality) in order to assess his/her fear of dogs at the behavioral level. Then, before the exploration of auditory–visual VEs, the participant became acquainted with the equipment and the navigation mode in a training immersion. The experimental exploration of two different auditory–visual VEs aimed at measuring fear in response to different sensory presentations of stimuli. Then, he/she was submitted a second time to a BAT with the same procedure as the first time. Finally, the participant completed several questionnaires and was asked by the experimenter to comment on his experience (debriefing). During the immersions in the different VEs, skin conductance was recorded.

All participants provided written informed consent prior to the experiment, which was approved by the Health Research Ethics Committee (CERES) of Paris Descartes University.

2.1. Participants

Participants were selected on the basis of their scores on a questionnaire exploring the fear of dogs (Viaud-Delmon *et al.*, 2008; see details in Section 2.5).

Twenty-two healthy volunteers (12 females; age: $M = 37.09$, $\mathrm{SD} = 13.78$) with normal or corrected to normal vision and audition were recruited to par-

Table 2.
Participants' characteristics

Variable	All participants	NoFear group	DogFear group
Number of individuals	$N = 21$	$n_{\text{NoFear}} = 10$	$n_{\text{DogFear}} = 11$
% of females[a]	52.38%	40.00%	63.64%
Age ($M \pm$ SD)[a]	36.00 ± 13.11	32.50 ± 12.06	39.18 ± 13.77
Trait anxiety score ($M \pm$ SD)[a]	41.29 ± 7.46	38.40 ± 6.83	43.91 ± 7.31
Dog fear score ($M \pm$ SD)[b]	12.95 ± 11.09	2.20 ± 1.32	22.73 ± 4.86

[a] Both groups were similar in terms of ratio of female (χ^2 test with Yates correction: $\chi^2_{(1)} = 0.42$, $p = 0.519$), age (Mann–Whitney test: $U = 38.00$, $p = 0.231$) and trait anxiety scores (Mann–Whitney test: $U = 32.50$, $p = 0.113$).

[b] The dog fear score was significantly different between groups (Mann–Whitney test: $U = 0.00$, $p < 0.001$).

ticipate in the study. None of them had a history of psychiatric disorder, neurological disorder or was under medical treatment. Twelve individuals (eight females; age: $M = 40.92$, SD $= 14.44$) had high dog fear scores and composed the DogFear group. The remaining ten individuals (four females; age: $M = 32.50$, SD $= 12.06$) had a low dog fear score and composed the NoFear group.

Only 21 among the 22 volunteers (see details in Table 2) participated in the second session of the experiment because one individual from the DogFear group broke his leg between the sessions.

2.2. *Virtual Reality Setup*

The experiment took place at INRIA in Sophia Antipolis. The immersive space was a BARCO iSpace, a four-sided, retro-projected cube with Infitec stereoscopic viewing (Fig. 1). Participants wore polarized glasses. The auditory scenes were presented through Sennheiser HD650 headphones and the sound stimuli were processed through binaural rendering using a non-individual Head Related Transfer Function (HRTF) of the LISTEN HRTF database (http://recherche.ircam.fr/equipes/salles/listen/) previously selected as best-fitting HRTF for a majority of participants to different experiments involving binaural rendering (see Moeck *et al.*, 2007; Sarlat *et al.*, 2006). The scenes had an ambient audio environment rendered through virtual ambisonic sources and binaural audio rendering. Head movements were tracked using an ART optical system so that visual stereo and 3D sounds were appropriately rendered with respect to the users' position and orientation. The participants were equipped with a wireless joystick to navigate in the VEs. With this device, they controlled both rotations and translations within the VEs.

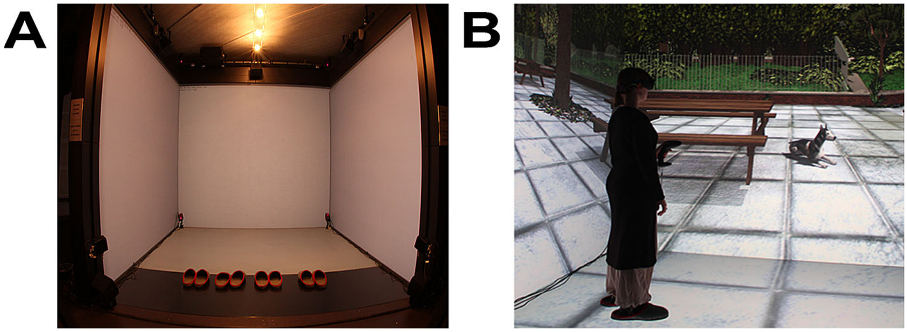

Figure 1. (A) Picture of the iSpace setup used in the study. (B) A participant, equipped with polarized glasses, headphones and a wireless joystick, standing within the iSpace during immersion in an auditory–visual VE. This figure is published in colour in the online version.

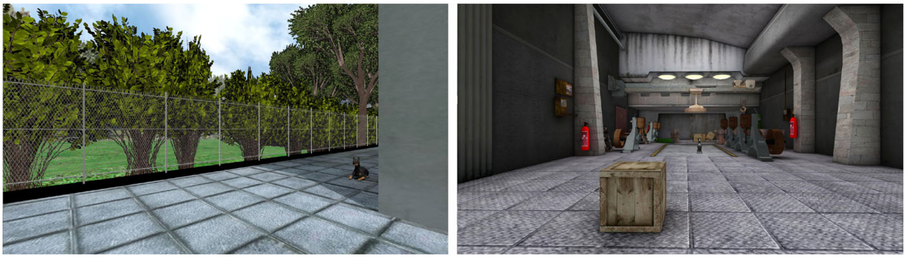

Figure 2. Pictures of the auditory–visual VEs used to measure the participants' fear when encountering virtual dogs. On the left, the outdoor garden scene and on the right, the indoor hangar scene. This figure is published in colour in the online version.

2.3. Virtual Environments

The VE used for the BAT was composed of a visual corridor and did not provide any auditory stimulation. The VE for training was a dog-free outdoor scene with trees and houses. Two different auditory–visual VEs were used to measure the participants' fear when encountering virtual dogs in different sensory conditions (Fig. 2). The first auditory–visual VE presented to participants was an outdoor garden scene composed of houses, trees and benches. The second auditory–visual VE was an indoor virtual scene in a large dark hangar, in which different pieces of industrial machinery were active. Auditory–visual VEs had an ambient audio environment composed of sounds of birds, of hustle and bustle sounds in the outdoor scene, and sounds of industrial machinery in the indoor scene.

2.4. Dog Stimuli

A Doberman model with three different textures was used (Fig. 3). The dog stimulus that was displayed during the BATs was a unimodal visual dog. In

Figure 3. Pictures of the virtual dog stimuli used in this study: the Doberman dog model with (from left to right) Malamute, Miniature Pinscher and Doberman texture. This figure is published in colour in the online version.

Table 3.
Dog stimuli and their presentation order in the auditory–visual VEs

1	Auditory static dog	Barking
2	Visual static dog	A dog lying
3	Auditory moving dog[a]	Looming and receding barking
4	Visual moving dog[a]	Dog standing up
5	Auditory–visual static dog	Dog lying down and growling
6	Auditory–visual moving dog[a]	Dog standing up and growling
7	Auditory–visual following dog[a]	Dog standing up, growling and following
8	Lower visual contrast[b]	Fog or dimming the light

[a] The dynamic stimuli were lying down and standing up when participants approached.
[b] Dog stimulus 8 was dog stimulus 7 with a lower visual contrast.

the auditory–visual VEs, the dog stimuli could be unimodal or bimodal, static or dynamic. Seven virtual dogs were displayed in a progressive manner during the exploration (see Table 3). There were a total of eight stimuli with the two last stimuli corresponding to the same virtual dog displayed with different visual contrasts.

2.5. Questionnaires and Interview Measures

The dog phobia questionnaire (Viaud-Delmon *et al.*, 2008) used to select participants consists of two sections. The first section asks four yes/no questions about reactions to dogs and the second section comprises 14 questions rated on a scale of 0 (no fear) to 3 (extreme fear), assessing fear in response to size of dog, activity level of dog, and physical restraint of dog (e.g. leash). The minimal score on this dog phobia questionnaire is 0 and the maximal one is 42. Two hundred and twenty-five individuals (98 females; age: $M = 31.71$, $SD = 11.40$) completed this questionnaire. A mean dog fear score ($M = 10.63$, $SD = 8.55$) was obtained, which served as a basis to select participants with high dog fear score (score $> M +$ SD) and low dog fear score (score $< M -$ SD) for the current experiment.

We used the State Trait Anxiety Inventory (STAI) (Spielberger *et al.*, 1983) to measure anxiety levels. Participants completed the trait version online several months before the experiment. The state portion of the STAI was used in the second session of the experiment, upon arrival at the laboratory as well as after completion of the total procedure. A 22-item cybersickness scale (Viaud-Delmon *et al.*, 2000) and the presence questionnaire from the I-group (Schubert *et al.*, 2001) were presented at the end of the immersions in the auditory–visual VEs.

Fear ratings were collected during immersion in both auditory–visual VEs as well as during the BATs using the Subjective Unit of Distress (SUD; Wolpe, 1973). SUD is a self-report measurement of fear level on a 0–100 point scale, which is widely used in behavioral research and therapy (e.g. Botella *et al.*, 1998; Emmelkamp *et al.*, 2001; Rothbaum *et al.*, 1995) and has been shown to correlate with several physiological measures of arousal (Thyer *et al.*, 1984).

2.6. Physiological Acquisitions

During immersions in the auditory–visual VEs, we monitored participants' SCL using two sensors that were attached to the palmar surface of the middle phalanges of the index and middle fingers of the non-dominant hand. A baseline was recorded for two minutes in the iSpace, before each immersion. Skin had been previously cleansed with alcohol. Participants were instructed to keep their hand relaxed and still during the recordings. Recordings were carried out by the wireless measurement device Captiv-L7000 (TEA, France) and sampled at 32 Hz.

2.7. Procedure

The participants completed the state portion of the STAI upon arrival. Then, they had to complete five immersions in virtual reality (BAT1, training, outdoor scene, indoor scene, BAT2).

Each participant was first invited to participate in the BAT1. During this immersion, the participant was standing at a precise spot on the extremity of a long corridor and a virtual unimodal visual dog was standing far (at 16.55 m) in front of him/her. The BAT was composed of 14 steps. The first step was for the participant to begin immersion and thus to face the dog for the first time. Then, at each of the next twelve steps, the virtual dog walked 1.25 m towards the participant, stopped and sat. For the final step, the participant had to approach the virtual dog by making a real step in the iSpace in order to put his/her face against the face of the virtual dog. At this point the participant could look at the dog from a 5-centimeter distance. At each of the 14 steps, he/she had to rate his/her anxiety level with SUDs. At each step the experimenter proposed stopping the test if the participant was feeling too anxious. If he/she felt ready, the next step was started. The BAT score scale was from 0 to 14 where 0 is refusal to begin immersion and 14 is putting one's face against the face of the virtual dog for more than five seconds.

Then, the participant went through a training immersion in order to become acquainted with the equipment and the navigation mode. During this training immersion, the experimenter interacted with the participant in order to assist him/her in his/her first navigation.

After training, the participant was immersed in the auditory–visual VEs, aiming to measure participant's reaction to the auditory and visual virtual dog stimuli. Each participant explored first the outdoor scene and then the indoor scene. He/she was instructed that there was a frog somewhere in the auditory–visual VEs and that his/her task was to explore them to find the frog. The frog was an auditory–visual object and could be both seen and heard. It was placed in the VEs so that participants could not find it before encountering all the dog stimuli. The participant was informed that he/she would encounter several dogs when completing his/her task. The sound spatialization played a major role in this case, as the participant could rely on the auditory information to locate both the dogs and the frog. Each participant explored the auditory–visual VEs freely. However, the scenarios were designed so that all participants had to take a certain path ensuring that virtual dogs were displayed in a progressive manner during the exploration, as described previously. The first six stimuli were displayed at fixed locations of the VEs while the last auditory–visual and dynamic dog followed the participants until they found the frog. During the exploration time where they were accompanied by this virtual dog, participants did not encounter any other dog stimulus. As a last step, we modified the visual contrast, by introducing fog in the outdoor scene and dimming the lights in the indoor scene. At each encounter with a dog stimulus, the participant had to rate his/her anxiety level with SUDs as well as when the visual contrast was modified.

After exposure in the auditory–visual VEs, the participant filled the presence questionnaire from the I-group and the cybersickness scale. Then, he/she participated in the BAT2. He/she also completed a second state portion of the STAI. Finally, a debriefing interview was conducted to collect feelings and impressions from the participant.

2.8. *Control Experiment: Assessment of Aversiveness of Barking* vs. *Growling*

For technical reasons (problems with lip synchronization), we had to use different dog sounds in the unimodal and bimodal conditions to ensure the coherence of the stimulations. In previous work evaluating dog stimuli, dog-fearful participants did not point out the type of the dog sound (barking or growling) as a factor having an impact on their fear (Suied *et al.*, 2013; Viaud-Delmon *et al.*, 2008). However, since the factors influencing the fear could be different among individuals, we tested the effect of this factor in our sample of participants. After the experiment, they had to complete a control test on-line. This test consisted in indicating the level of fear they experienced when hearing each sound (barking and growling), by using SUDs. The sounds were displayed for eleven seconds and the presentation order was counterbalanced between subjects inside both NoFear and DogFear groups.

2.9. Data Analyses

Differences between groups were evaluated using two-tailed non-parametric Mann–Whitney U tests. Comparisons within each group were performed using the two-tailed non-parametric Wilcoxon T test for matched samples.

Two dog-fearful individuals did not complete the protocol because of strong manifestations of the autonomic nervous system related to virtual reality (cybersickness). The analyses were conducted on the 19 individuals ($n_{\mathrm{NoFear}} = 10$; $n_{\mathrm{DogFear}} = 9$), who participated in each of the five immersions, completing the second session.

2.9.1. Questionnaire Measures

One pre-immersion state anxiety score was lost. We compared participants' pre- and post-immersion state anxiety scores, between and within groups ($n_{\mathrm{NoFear}} = 9$; $n_{\mathrm{DogFear}} = 9$). We compared cybersickness and presence scores between groups ($n_{\mathrm{NoFear}} = 10$; $n_{\mathrm{DogFear}} = 9$).

2.9.2. Behavioral Assessment of Dog Fear (BATs)

We compared participants' scores and mean SUDs per step on the BAT1 and BAT2, between and within NoFear and DogFear groups. For each participant, we summed the SUDs they reported at each step of BATs. We divided this sum by the number of steps the participant managed to go through (score) in order to obtain a mean SUD per step for each of the BATs. Within each group, we also investigated the modifications of fear level from step to step by conducting multiple comparisons using a two-tailed non-parametric Wilcoxon T test for matched samples. In order to address possible α error accumulation, p-values are given as calculated, for interpretation of results classical Bonferroni correction for multiple testing was considered.

2.9.3. Sensory Modality and Fear in the Auditory–Visual VEs

First, we compared the mean level of fear during immersion in the auditory–visual VEs between groups. For each participant, we averaged all SUDs reported in the auditory–visual VEs and compared the resulting mean SUDs between the NoFear and DogFear groups.

Then, we tested the effect of the VE (Outdoor/Indoor) on fear. Within each group, we averaged SUDs reported in the outdoor VE on one hand and SUDs reported in the indoor VE on the other hand and compared them. We also tested the effect of visual contrast on fear by comparing SUDs in response to the seventh and eighth stimuli.

In order to compare the fear evoked by unimodal and bimodal stimuli in the auditory–visual VEs, we calculated the mean SUDs according to the sensory modality in which the dogs were presented. Among the SUDs reported during the immersion in the VEs, we averaged the SUDs collected in response to the four unimodal dog stimuli on the one hand and to the first two bimodal dog stimuli on the other hand. In the bimodal condition, the average of SUDs did not include the data in response to the third bimodal dog stimulus. This stimulus, which followed the participant, had no counterpart in the unimodal condition and increased the mean SUDs in the bimodal condition if included.

We also calculated the sum of the mean SUD in response to the visual stimuli and of the mean SUD in the auditory condition in the whole sample. We compared this sum to the mean SUD in response to the first two auditory–visual stimuli. We verified the effect of the order of stimuli presentation. We averaged the SUDs in response to the four unimodal stimuli in each of the VEs and compared the resulting mean SUDs.

2.9.4. Sensory Modality and Fear-Related Physiological Arousal in the Auditory–Visual VEs

Seven participants (five NoFear and two DogFear) were excluded from the analysis because of missing data and/or noisy signal due to the limitations of the space and the equipment (the recording PC had to be outside the iSpace, and the walls of the iSpace interfered with transmission of the signal). Given the few remaining participants in each group, we analyzed the data globally without taking account of groups ($N = 12$).

Skin conductance data were analyzed using the Matlab analysis software Ledalab (V3.4.1) (Benedek and Kaernbach, 2010). First, artifacts were manually detected and rejected. Then, the Ledalab's Continuous Decomposition Analysis was run, optimizing the fit and reducing the error of the model. This method returns the SCL as a continuous measure of tonic electrodermal activity and the phasic driver as a continuous measure of phasic electrodermal activity. For each participant, we extracted mean SCL during immersion (SCLi) and during the baseline (SCLb). Then, we calculated the normalized mean SCL during immersion (nSCL) as follow: $\text{nSCL} = ((\text{SCLi} - \text{SCLb})/\text{SCLb})$.

We first tested the effect of the VE (Outdoor/Indoor) on participants' physiological arousal by comparing nSCL during the Outdoor VE to nSCL during the Indoor VE. Then, we compared nSCL during unimodal and during bimodal presentation of dog stimuli. We verified the effect of the order of stimulus presentation by comparing nSCL during unimodal presentation of dog stimuli between the two auditory–visual VEs.

2.9.5. *Control Experiment: Assessment of Aversiveness of Barking* vs. *Growling*

One participant from the NoFear group did not complete the control experiment. The analyses were conducted on the 18 remaining participants ($n_{\text{NoFear}} = 9$; $n_{\text{DogFear}} = 9$) who completed both the protocol and the control experiment. Within each group, we compared the SUDs in response to the barking sound to the SUDs in response to the growling sound.

3. Results

Two individuals from the DogFear group did not complete the protocol because of strong cybersickness. The first one stopped during training and the second one during immersion in the outdoor scene. Their scores on the cybersickness scale were respectively 18 and 17. All non-fearful individuals participated in each of the five immersions.

Our analyses did not reveal any sex differences.

3.1. *Questionnaire Measures*

The state anxiety scores of all non-fearful participants decreased after the immersions. The NoFear group scores were significantly lower after the immersions compared to before ($T = 0.00$, $p = 0.008$). In the DogFear group, four individuals had a state anxiety score that was lower after than before the immersions, four had a higher score after the immersions and the last one had the same score in both assessments (see Table 4). In this group, the mean state anxiety score was not significantly different between pre- and post-immersion ($T = 16.50$, $p = 0.834$).

There was no difference in state anxiety scores (State anxiety 1: $U = 40.50$, $p = 1.000$; State anxiety 2: $U = 33.50$, $p = 0.348$), cybersickness scores ($U = 42.50$, $p = 0.838$) or presence scores ($U = 27.00$, $p = 0.142$) between the two groups.

3.2. *Measures During BATs*

Among the 19 participants who completed the study, 16 reached the final step in both BAT immersions (BAT1 and BAT2) and thus obtained maximal scores. The other three individuals did not manage to get to the end of either BAT immersion because of anxiety. They all belonged to the DogFear group. The BAT1 scores of the NoFear group ($M_{\text{NoFear}} = 14.00$, $\text{SD}_{\text{NoFear}} = 0.00$) and the BAT1 scores of the DogFear group ($M_{\text{DogFear}} = 13.56$, $\text{SD}_{\text{DogFear}} = 0.73$) were not significantly different ($U = 30.00$, $p = 0.221$). The BAT2 scores of the NoFear group ($M_{\text{NoFear}} = 14.00$, $\text{SD}_{\text{NoFear}} = 0.00$) and the BAT2 scores of the DogFear group ($M_{\text{DogFear}} = 13.44$, $\text{SD}_{\text{DogFear}} = 1.01$) were also not

Table 4.
Individual questionnaire measures

ID	State anxiety 1	State anxiety 2	Cybersickness	Presence
Possible range	[20–80]		[0–88]	[0–84]
NoFear group				
NF-1	/	26	6	33
NF-2	24	21	1	43
NF-3	31	25	6	44
NF-4	23	22	16	57
NF-5	28	23	0	43
NF-6	21	20	7	40
NF-7	24	22	5	52
NF-8	26	23	5	44
NF-9	33	23	3	23
NF-10	32	25	3	40
$M \pm$ SD	26.89 ± 4.31	23.00 ± 1.89	5.20 ± 4.42	41.90 ± 9.34
DogFear group				
DF-1	32	39	0	50
DF-2	31	55	20	43
DF-3	24	33	1	60
DF-4	27	24	10	27
DF-5	31	24	8	49
DF-6	38	27	25	46
DF-7	20	20	0	68
DF-8	20	21	0	66
DF-9	24	20	1	40
$M \pm$ SD	27.44 ± 6.04	29.22 ± 11.57	7.22 ± 9.50	49.89 ± 13.11

significantly different ($U = 30.00$, $p = 0.221$). In both groups, there was no significant difference between BAT1 and BAT2 scores.

In both BATs, the mean SUD per step was higher for the DogFear group (BAT1: $M_{\text{DogFear}} = 17.78$, $\text{SD}_{\text{DogFear}} = 11.34$; BAT2: $M_{\text{DogFear}} = 15.95$, $\text{SD}_{\text{DogFear}} = 14.06$) compared to the NoFear group (BAT1: $M_{\text{NoFear}} = 2.14$, $\text{SD}_{\text{NoFear}} = 2.28$; BAT2: $M_{\text{NoFear}} = 0.79$, $\text{SD}_{\text{NoFear}} = 1.32$; BAT1: $U = 2.00$, $p < 0.001$; BAT2: $U = 5.00$, $p = 0.001$). The mean SUD per step was not different between BAT1 and BAT2 for any of the groups (NoFear group: $T = 4.00$, $p = 0.091$; DogFear group: $T = 16.00$, $p = 0.441$).

In the DogFear group, there was a global increase of SUDs in both BATs, as the virtual dog got closer (Fig. 4). The Wilcoxon test revealed a significant increase of SUDs between step 11 and step 12 ($T = 0.00$, $p = 0.018$; $n_{\text{DogFear}} = 9$) and between step 12 and step 13 ($T = 0.00$, $p = 0.028$; $n_{\text{DogFear}} = 8$) in BAT1. The transition between step 11 and step 12 corresponded to the dog

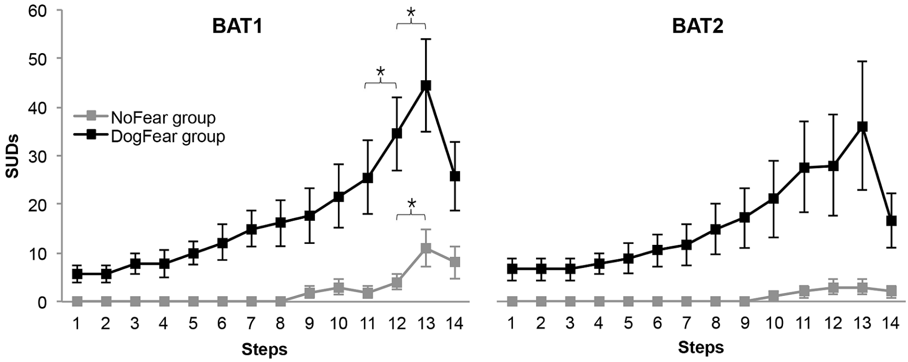

Figure 4. Mean reported fear (mean SUDs ± SEM) of the NoFear group (grey squares) and the DogFear group (black squares) at each of the 14 steps during BATs. The responses collected during BAT1 are presented on the left and the responses collected during BAT2 are presented on the right. In the NoFear group, Wilcoxon tests revealed a significant increase of fear between steps 12 and 13 of BAT1. In the DogFear group, fear increased globally in both BATs and Wilcoxon tests revealed significant increases of fear between steps 11 and 12 and between steps 12 and 13 in BAT1. Neither groups showed any increase of fear between steps in BAT2.

approaching from 4.05 m to 2.80 m distance from participants. The transition between step 12 and step 13 corresponded to the dog approaching from 2.80 m to 1.55 m distance from participants. In BAT2, the Wilcoxon test did not indicate any significant increase of SUDs between steps in the DogFear group.

Globally, there was no increase of SUDs during BATs in the NoFear group (see Fig. 4). However, the Wilcoxon test indicated a significant increase of SUDs between step 12 and step 13 in BAT1 ($T = 0.00$, $p = 0.043$, $n_{\text{NoFear}} = 10$). In BAT2, the Wilcoxon test did not reveal any significant increase of SUDs between steps in the NoFear group.

Within each group, we conducted 13 comparisons. With the Bonferroni correction (corrected p-value $= 0.004$), we did not find any significant difference of SUDs between steps in either of the BATs.

3.3. Measures During Immersion in the Auditory–Visual VEs

3.3.1. Sensory Modality and Fear: Subjective Units of Distress (SUDs)

The DogFear group reported higher SUDs ($M_{\text{DogFear}} = 27.49$, $\text{SD}_{\text{DogFear}} = 13.73$) compared to the NoFear group ($M_{\text{NoFear}} = 4.71$, $\text{SD}_{\text{NoFear}} = 2.71$) in the auditory–visual VEs ($U = 0.00$, $p < 0.001$).

Within each group, the two auditory–visual VEs provoked the same level of fear: SUDs were not significantly different between VEs in the NoFear group ($T = 11.00$, $p = 0.173$, $M_{\text{NoFear/Outdoor}} = 5.50$, $\text{SD}_{\text{NoFear/Outdoor}} = 3.55$; $M_{\text{NoFear/Indoor}} = 3.90$, $\text{SD}_{\text{NoFear/Indoor}} = 3.87$) or in the DogFear group ($T = 18.00$, $p = 0.594$, $M_{\text{DogFear/Outdoor}} = 26.34$, $\text{SD}_{\text{DogFear/Outdoor}} = 17.43$;

Table 5.
SUDs ($M \pm$ SD) in response to the stimuli during immersion in the VEs

Stimulus	NoFear group	DogFear group
Auditory static dog	0.00 ± 0.00	10.00 ± 5.00
Visual static dog	0.50 ± 1.58	13.06 ± 9.82
Auditory moving dog	0.00 ± 0.00	14.17 ± 10.16
Visual moving dog	2.50 ± 2.64	20.00 ± 15.00
Auditory–visual static dog	6.50 ± 7.47	38.06 ± 19.11
Auditory–visual moving dog	9.50 ± 6.85	39.17 ± 22.64
Auditory–visual following dog	9.50 ± 6.85	46.88 ± 30.47
Lower visual contrast	8.00 ± 7.89	44.29 ± 31.01

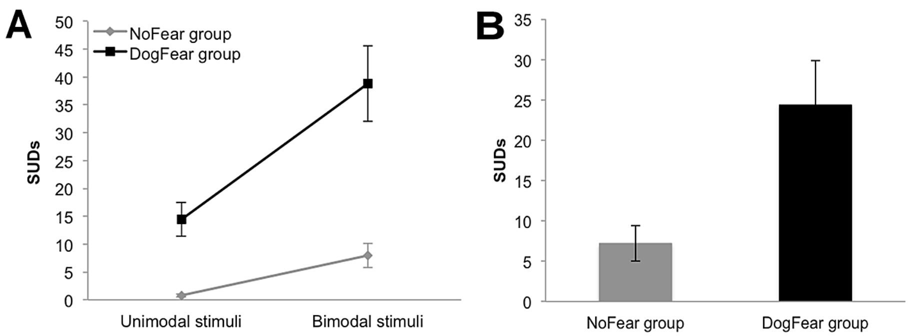

Figure 5. (A) Mean reported fear (mean SUDs ± SEM) of the NoFear (grey diamonds) and DogFear group (black squares) in the auditory–visual VEs according to the sensory modality in which the dogs were presented. The SUDs reported in response to the auditory static, the visual static, the auditory moving and the visual moving dog stimuli were averaged for the unimodal condition. The SUDs in response to the auditory–visual static and the auditory–visual moving dog stimuli were averaged for the bimodal condition. In both groups, the experience of fear was higher in response to bimodal compared to unimodal stimuli. (B) Mean increase of reported fear in the bimodal condition compared to the unimodal one (mean difference between SUDs in response to bimodal and unimodal stimuli ± SEM) in each group. The increase of fear is greater in the DogFear group (black bar) than in the NoFear group (grey bar).

$M_{\text{DogFear/Indoor}} = 28.64$, $\text{SD}_{\text{DogFear/Indoor}} = 10.17$). Since there was no effect of VE, we averaged SUDs from both VEs (see Table 5). There was no significant difference of fear level between the seventh and the eighth stimuli in the DogFear group ($T = 4.00$, $p = 0.173$) or in the NoFear group ($T = 6.00$, $p = 0.345$).

As Fig. 5 shows, SUDs were higher for bimodal stimuli compared to unimodal ones for both groups (NoFear group: $T = 0.00$, $p = 0.008$; DogFear group: $T = 0.00$, $p = 0.008$). Moreover, the increase of SUDs for bimodal stimuli was higher in the DogFear group compared to the NoFear group ($U = 5.00$, $p = 0.001$).

In the whole sample, the mean SUD in the bimodal condition was higher than the sum of the mean SUDs from each unimodal condition ($T = 18.50$, $p = 0.006$; $M_{\text{Bimodal}} = 22.59$, $\text{SD}_{\text{Bimodal}} = 21.35$; $M_{\text{Sum Unimodal}} = 14.96$, $\text{SD}_{\text{Sum Unimodal}} = 19.25$). The mean SUD in response to the unimodal stimuli in the Indoor VE was not different from the mean SUD in response to unimodal stimuli in the Outdoor VE ($T = 25.00$, $p = 0.272$; $M_{\text{Unimodal/Outdoor}} = 6.32$, $\text{SD}_{\text{Unimodal/Outdoor}} = 8.91$; $M_{\text{Unimodal/Indoor}} = 8.42$, $\text{SD}_{\text{Unimodal/Indoor}} = 10.55$).

3.3.2. Sensory Modality and Fear-Related Physiological Arousal: Skin Conductance Level (SCL)

The two auditory–visual VEs provoked the same level of nSCL ($T = 24.00$, $p = 0.239$, $M_{\text{Outdoor}} = 0.088$, $\text{SD}_{\text{Outdoor}} = 0.111$; $M_{\text{Indoor}} = 0.027$, $\text{SD}_{\text{Indoor}} = 0.140$; $N = 12$). The nSCL was lower during unimodal stimulations ($M_{\text{unimodal}} = 0.038$, $\text{SD}_{\text{unimodal}} = 0.117$, $N = 12$) compared to bimodal stimulations ($M_{\text{bimodal}} = 0.077$, $\text{SD}_{\text{bimodal}} = 0.099$, $N = 12$) in the VEs ($T = 11.00$, $p = 0.028$). The nSCL during unimodal stimulations in the Indoor VE was not different from the nSCL during unimodal stimulations in the Outdoor VE ($T = 20.00$, $p = 0.136$, $N = 12$).

3.4. Control Experiment: Assessment of Aversiveness of Barking vs. *Growling*

The DogFear group reported higher SUDs in response to the growling sound compared to the barking sound ($T = 0.00$, $p = 0.008$; $M_{\text{DogFear/barking}} = 15.00$, $\text{SD}_{\text{DogFear/barking}} = 10.31$; $M_{\text{DogFear/growling}} = 38.33$, $\text{SD}_{\text{DogFear/growling}} = 21.65$). In the NoFear group, there was no significant difference between the SUDs in response to the barking and to the growling sound ($T = 3.00$, $p = 0.465$; $M_{\text{NoFear/barking}} = 9.44$, $\text{SD}_{\text{NoFear/barking}} = 21.13$; $M_{\text{NoFear/growling}} = 12.22$, $\text{SD}_{\text{NoFear/growling}} = 16.60$).

4. Discussion

The goal of this study was to determine whether multisensory presentation of aversive stimuli has an influence on the conscious experience of fear. Our study shows that the auditory–visual presentation of aversive stimuli modulates affect. Auditory–visual aversive stimuli increase the conscious experience of fear.

We exploited the unique advantages of virtual reality concerning the manipulation of multimodal stimuli inputs and their naturalistic display (Bohil *et al.*, 2011). We compared the experience of fear (SUDs) induced by unimodal and bimodal dog stimuli in healthy participants. We modulated the fear evoked by dog stimuli by recruiting two categories of participants: dog-fearful participants (DogFear group) and non-fearful participants (NoFear group). During the BATs, the NoFear group did not report any global fear while the DogFear group reported an increasing fear, as the unimodal dog got closer to them. Moreover, while each non-fearful participant completed the test, three dog-fearful participants did not complete the test. These results confirm the fact that at the behavioral and at the subjective level the DogFear group considers dogs as aversive while the NoFear group considers them as non-aversive. This fact validates the use of these two groups to modulate the fear in response to our dog stimuli.

A narrower analysis of the BATs offers further interesting results. The results of this analysis did not resist the Bonferroni correction; consequently the findings are discussed as hypothesis generating rather than as confirmatory (Streiner and Norman, 2011).

In BAT1, we observed that both NoFear and DogFear participants' SUDs increased when the dog approached to a relatively small distance from them. In dog-fearful participants, this enhanced fear would be consistent with aversive stimuli representing a higher threat when intruding participants' near space. In non-fearful participants, this would be consistent with fear-relevant stimuli turning to aversive stimuli when intruding the near space. The limit distance was higher for the dog-fearful participants (between 4.05 and 2.80 m) than for the non-fearful participants (between 2.80 and 1.55 m). This suggests that the dog-fear level may influence the distance perception between themselves and a dog stimulus. Recent studies have also found an effect of the level of fear on distance perception in height, claustrophobic, snake and spider fear (Clerkin *et al.*, 2009; Lourenco *et al.*, 2011; Vagnoni *et al.*, 2012). Our results fit with these findings and extend them by suggesting that the level of dog fear may impact distance perception.

Surprisingly, while participants' fear in BAT1 increased significantly when the distance to the dog got smaller, the SUDs did not increase at the final step. At this final step, it was not the dog who approached the participant but the participant who approached the dog. It seems that this configuration is less threatening. This result suggests an effect of objective stimulus control and subjective feelings of controllability on the experience of fear. This is in line with previous observations that perceiving control over aversive events influences how we experience them (Buetti and Lleras, 2012; Leotti *et al.*, 2010).

The narrow analysis of BAT2 showed different results than BAT1. After the immersion in the auditory–visual VEs, the dog–participant distance evoking an enhancement of fear in BAT2 decreased in both groups. The limit distance was shorter than 1.55 m in both groups. The virtual unimodal dog could approach closer to participants before evoking an enhancement of fear. This suggests that one immersion in our auditory–visual VEs containing virtual dogs reduced the fear of dogs. This would not be a surprise since our procedure is very similar to protocols of virtual reality-based exposure therapy, which are used for the treatment of anxiety disorders (e.g. Botella *et al.*, 1998; Emmelkamp *et al.*, 2001; Garcia-Palacios *et al.*, 2002; Riva, 2005; Rothbaum *et al.*, 1995; Wald and Taylor, 2001).

In our protocol, during the immersion in the auditory–visual VEs, we presented the supposedly less fearful (unimodal) stimuli before the supposedly most fearful (bimodal) stimuli to avoid saturation effects and access the experience of fear in both conditions. In both the DogFear and NoFear groups, we observed higher SUDs in the bimodal condition relative to the unimodal condition. The auditory–visual aversive stimuli evoked an increased experience of fear. A similar effect has been put forward on aesthetic experience in response to musical performances (Vines *et al.*, 2006, 2011). The visual inputs modulate the judgment of tension in the performance and the Likert-scale ratings of the intensity of the positive emotion experienced during the performance. Baumgartner *et al.* also showed a similar effect on emotional experience in response to positive and negative stimuli (Baumgartner *et al.*, 2006). They combined International Affective Picture System (IAPS) pictures and music excerpts and demonstrated that affective music stimuli enhance the arousal experience in response to affective pictures. By using and manipulating the sensory inputs of a natural multisensory stimulus, our findings show that the multisensory presentation of stimuli enhances the experience of emotion.

In the entire population, we also observed a higher SCL during bimodal presentation compared to SCL during unimodal presentation of the dog stimuli. This increase of physiological arousal is in line with the findings on positive musical performances (Chapados and Levitin, 2008), and suggests that multisensory stimulation would enhance motivational (appetitive or defensive) engagement by increasing non-conscious emotion (Bradley *et al.*, 2001; Kreibig, 2010).

The limitations of these results are closely linked to their strengths. First, our protocol did not allow controlling the potential effect of the presentation order. However, the SUDs and the SCL in the unimodal condition are not different between the outdoor and the indoor VEs despite the fact that participants encountered the unimodal stimuli in the indoor VE after encountering the bimodal stimuli in the outdoor VE. Consequently, the increased experience of fear and physiological arousal in the bimodal condition cannot be attributed to the presentation order.

Second, we had to deal with the technological challenges of virtual reality, which constrained us to use different dog sounds in the unimodal and bimodal conditions. We used a barking sound as unimodal auditory stimulus and a growling sound in the auditory–visual condition to avoid problems with lip synchronization. In the control experiment, the DogFear group reported greater fear in response to the growling sound compared to the barking sound. One may wonder if the effect of multisensory presentation in this group is, in fact, completely linked to the use of the growling sound in the bimodal condition. However, the multisensory presentation also enhances the fear in the NoFear group although they did not report a different level of fear between both sounds in the control experiment. It is therefore likely that in the dog-fearful group, the enhancement of fear in the bimodal condition is due to both the multisensory presentation and the use of the growling sound rather than only the use of the growling sound.

There are several possibilities as to how auditory–visual presentation might influence the conscious experience of emotion. The present effect could be explained in terms of arousal. Testing this hypothesis in an ecological protocol such as ours is difficult. We indeed cannot repeat and mix the different fearful stimuli, since we need a gradation of stimulus aversiveness. However, the modification of visual contrast, which is a manipulation of arousal, did not create an effect comparable to the effect of auditory–visual presentation.

Second, the increased fear in the bimodal condition may be linked to multisensory processes. The effect of the auditory information and the effect of the visual information on the experience of emotion could be independent. In this case, the enhancement of fear would be linked to an additive effect (Stein and Meredith, 1993; Stein and Stanford, 2008). Alternatively, the inputs coming from both senses could interact to further enhance the experience of fear. In that case, the enhancement of fear would be linked to a cross-modal potentiation (a subadditive or superadditive effect) (Stein and Meredith, 1993; Stein and Stanford, 2008). The evaluation stage of multisensory processing of affective face–voice pairs has been investigated with fMRI, PET and MEG techniques. Cerebral activation around the superior temporal sulcus (STS) has been found to be higher in response to auditory–visual affective stimuli than to the conjunction or addition of auditory and visual presentation of the stimuli (see Ethofer *et al.*, 2006 as a review; Hagan *et al.*, 2009; Kreifelts *et al.*, 2007; Pourtois *et al.*, 2005; Robins *et al.*, 2009). These results suggest interactions between the effects of the sensory inputs. Concerning the stage of feeling, the method used in our study did not allow investigating this question. However, the fear reported by participants in the auditory–visual condition was significantly higher than the sum of the fear reported in the auditory and the visual conditions. Although our paradigm cannot disentangle the two hypotheses, our data are rather in favor of the cross-modal potentiation hypothesis.

Besides, the effect of multisensory presentation on fear was different between groups. The increase of fear between the unimodal and the bimodal conditions was greater in dog-fearful participants relative to non-fearful participants. This effect could be linked to the growling sound evoking a greater fear in the DogFear compared to the NoFear group. It could also be accounted for by an influence of the dog fear level on multisensory processes since behavioral results on both human and animal models has suggested that anxiety impacts multisensory integration (Koizumi *et al.*, 2011; Viaud-Delmon *et al.*, 2011).

5. Conclusion

In spite of its limitations due to the use of an ecological paradigm using virtual reality, our study suggests that, beyond the facilitation of emotional judgment (Collignon *et al.*, 2008; Dolan *et al.*, 2001; Föcker *et al.*, 2011; Kreifelts *et al.*, 2007; Massaro and Egan, 1996), the multisensory presentation of affective stimuli enhances the conscious experience of emotion. This finding could be of great interest for the treatment of phobias. It indicates indeed that in order to completely address the disrupted affective processing, treatments for phobia should implicate and manipulate multisensory presentation of feared situations. Future investigations should focus on whether the enhancement of the experience of emotion in response to multisensory affective stimuli is due to an additive effect, a cross-modal potentiation or a simple arousal effect.

Acknowledgements

This research was supported by the EU FP7-ICT-2011-7 project VERVE (http://www.verveconsortium.eu/), grant nr. 288910. We are grateful to Prof. Philippe Robert and to Julie Piano for their help with the clinical evaluations. We thank Brigitte Trousse for the use of the skin-conductance measurement equipment, which was supported by FocusLab platform (PACA CPER telius) of the ICT Usage lab (EnOLL living Lab since 2006). We thank Stéphanie Dubal and Jane Mason for their comments on the manuscript.

References

Baumgartner, T., Lutz, K., Schmidt, C. F. and Jäncke, L. (2006). The emotional power of music: How music enhances the feeling of affective pictures, *Brain Res.* **1075**, 151–164. DOI:10.1016/j.brainres.2005.12.065

Benedek, M. and Kaernbach, C. (2010). A continuous measure of phasic electrodermal activity, *J. Neurosci. Meth.* **190**, 80–91. DOI:10.1016/j.jneumeth.2010.04.028

Bohil, C. J., Alicea, B. and Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy, *Nat. Rev. Neurosci.* **12**, 752–762. DOI:10.1038/nrn3122

Botella, C., Baños, R. M., Perpiñá, C., Villa, H., Alcañiz, M. and Rey, A. (1998). Virtual reality treatment of claustrophobia: a case report, *Behav. Res. Ther.* **36**, 239–246.

Bradley, M. M., Codispoti, M., Cuthbert, B. N. and Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing, *Emotion* **1**, 276–298. DOI:10.1037/1528-3542.1.3.276

Buetti, S. and Lleras, A. (2012). Perceiving control over aversive and fearful events can alter how we experience those events: An investigation of time perception in spider-fearful individuals, *Front. Psychol.* **3**, 337. DOI:10.3389/fpsyg.2012.00337

Chapados, C. and Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: Physiological correlates, *Cognition* **108**, 639–651. DOI:10.1016/j.cognition.2008.05.008

Chen, Y., Edgar, J. C., Holroyd, T., Dammers, J., Thönnessen, H., Roberts, T. P. L. and Mathiak, K. (2010). Neuromagnetic oscillations to emotional faces and prosody, *Eur. J. Neurosci.* **31**, 1818–1827. DOI:10.1111/j.1460-9568.2010.07203.x

Clerkin, E. M., Cody, M. W., Stefanucci, J. K., Proffitt, D. R. and Teachman, B. A. (2009). Imagery and fear influence height perception, *J. Anxiety Disord.* **23**, 381–386. DOI:10.1016/j.janxdis.2008.12.002

Collignon, O., Girard, S., Gosselin, F., Roy, S., Saint-Amour, D., Lassonde, M. and Lepore, F. (2008). Audio-visual integration of emotion expression, *Brain Res.* **1242**, 126–135. DOI:10.1016/j.brainres.2008.04.023

Damasio, A. R. (1998). Emotion in the perspective of an integrated nervous system, *Brain Res. Brain Res. Rev.* **26**, 83–86.

De Gelder, B., Böcker, K. B., Tuomainen, J., Hensen, M. and Vroomen, J. (1999). The combined perception of emotion from voice and face: Early interaction revealed by human electric brain responses, *Neurosci. Lett.* **260**, 133–136.

De Gelder, B., Pourtois, G. and Weiskrantz, L. (2002). Fear recognition in the voice is modulated by unconsciously recognized facial expressions but not by unconsciously recognized affective pictures, *Proc. Natl Acad. Sci. USA* **99**, 4121–4126. DOI:10.1073/pnas.062018499

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

De Gelder, B., Vroomen, J., De Jong, S. J., Masthoff, E. D., Trompenaars, F. J. and Hodiamont, P. (2005). Multisensory integration of emotional faces and voices in schizophrenics, *Schizophr. Res.* **72**, 195–203. DOI:10.1016/j.schres.2004.02.013

De Jong, J. J., Hodiamont, P. P. G. and De Gelder, B. (2010). Modality-specific attention and multisensory integration of emotions in schizophrenia: Reduced regulatory effects, *Schizophr. Res.* **122**, 136–143. DOI:10.1016/j.schres.2010.04.010

De Jong, J. J., Hodiamont, P. P. G., Van den Stock, J. and De Gelder, B. (2009). Audiovisual emotion recognition in schizophrenia: Reduced integration of facial and vocal affect, *Schizophr. Res.* **107**, 286–293. DOI:10.1016/j.schres.2008.10.001

Dolan, R. J., Morris, J. S. and De Gelder, B. (2001). Crossmodal binding of fear in voice and face, *Proc. Natl Acad. Sci. USA* **98**, 10006–10010. DOI:10.1073/pnas.171288598

Emmelkamp, P. M., Bruynzeel, M., Drost, L. and Van der Mast, C. A. (2001). Virtual reality treatment in acrophobia: A comparison with exposure *in vivo*, *Cyberpsychol. Behav.* **4**, 335–339.

Ethofer, T., Pourtois, G. and Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain, *Progr. Brain Res.* **156**, 345–361. DOI:10.1016/S0079-6123(06)56019-4

Föcker, J., Gondan, M. and Röder, B. (2011). Preattentive processing of audio-visual emotional signals, *Acta Psychol.* **137**, 36–47. DOI:10.1016/j.actpsy.2011.02.004

Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness, T. A. and Botella, C. (2002). Virtual reality in the treatment of spider phobia: a controlled study, *Behav. Res. Ther.* **40**, 983–993.

Garrett, A. S. and Maddock, R. J. (2001). Time course of the subjective emotional response to aversive pictures: Relevance to fMRI studies, *Psychiat. Res.* **108**, 39–48.

Hagan, C. C., Woods, W., Johnson, S., Calder, A., Green, G. G. R. and Young, A. W. (2009). MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus, *Proc. Natl Acad. Sci. USA* **106**, 20010–20015. DOI:10.1073/pnas.0905792106

Jessen, S. and Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions, *NeuroImage* **58**, 665–674. DOI:10.1016/j.neuroimage.2011.06.035

Koizumi, A., Tanaka, A., Imai, H., Hiramatsu, S., Hiramoto, E., Sato, T. and De Gelder, B. (2011). The effects of anxiety on the interpretation of emotion in the face–voice pairs, *Exp. Brain Res.* **213**, 275–282. DOI:10.1007/s00221-011-2668-1

Kreibig, S. D. (2010). Autonomic nervous system activity in emotion: a review, *Biol. Psychol.* **84**, 394–421. DOI:10.1016/j.biopsycho.2010.03.010

Kreifelts, B., Ethofer, T., Grodd, W., Erb, M. and Wildgruber, D. (2007). Audiovisual integration of emotional signals in voice and face: An event-related fMRI study, *NeuroImage* **37**, 1445–11456. DOI:10.1016/j.neuroimage.2007.06.020

Leotti, L. A., Iyengar, S. S. and Ochsner, K. N. (2010). Born to choose: The origins and value of the need for control, *Trends Cogn. Sci.* **14**, 457–463. DOI:10.1016/j.tics.2010.08.001

Lourenco, S. F., Longo, M. R. and Pathman, T. (2011). Near space and its relation to claustrophobic fear, *Cognition* **119**, 448–453. DOI:10.1016/j.cognition.2011.02.009

Magnée, M., De Gelder, B., Van Engeland, H. and Kemner, C. (2007). Facial electromyographic responses to emotional information from faces and voices in individuals with pervasive developmental disorder, *J. Child Psychol. Psychiat.* **48**, 1122–1130. DOI:10.1111/j.1469-7610.2007.01779.x

Magnée, M., De Gelder, B., Van Engeland, H. and Kemner, C. (2008). Atypical processing of fearful face–voice pairs in Pervasive Developmental Disorder: an ERP study, *Clin. Neurophysiol.* **119**, 2004–2010. DOI:10.1016/j.clinph.2008.05.005

Magnée, M., De Gelder, B., Van Engeland, H. and Kemner, C. (2011). Multisensory integration and attention in autism spectrum disorder: Evidence from event-related potentials, *PloS One* **6**, e24196. DOI:10.1371/journal.pone.0024196

Massaro, D. and Egan, P. (1996). Perceiving affect from the voice and the face, *Psychon. B Rev.* **3**, 215–221.

Maurage, P., Philippot, P., Joassin, F., Pauwels, L., Pham, T., Prieto, E. A., Palmero-Soler, E., Zanow, F. and Campanella, S. (2008). The auditory–visual integration of anger is impaired in alcoholism: An event-related potentials study, *J. Psychiatr. Neurosci.* **33**, 111–122.

Moeck, T., Bonneel, N., Tsingos, N., Drettakis, G., Viaud-Delmon, I. and Alloza, D. (2007). Progressive Perceptual audio rendering of complex scenes, in: *Proceedings of the 2007 Symposium on Interactive 3D Graphics and Games*, pp. 189–196. ACM, New York, NY, USA.

Mühlberger, A., Sperber, M., Wieser, M. J. and Pauli, P. (2008). A Virtual Reality Behavior Avoidance Test (VR-BAT) for the assessment of spider phobia, *J. CyberTher. Rehabil.* **1**, 147–158.

Müller, V. I., Cieslik, E. C., Turetsky, B. I. and Eickhoff, S. B. (2012). Crossmodal interactions in audiovisual emotion processing, *NeuroImage* **60**, 553–561. DOI:10.1016/j.neuroimage.2011.12.007

Nesse, R. M., Curtis, G. C., Brown, G. M. and Rubin, R. T. (1980). Anxiety induced by flooding therapy for phobias does not elicit prolactin secretory response, *Psychosom. Med.* **42**, 25–31.

Phillips, M. L., Drevets, W. C., Rauch, S. L. and Lane, R. (2003). Neurobiology of emotion perception I: The neural basis of normal emotion perception, *Biol. Psychiat.* **54**, 504–514. DOI:10.1016/S0006-3223(03)00168-9

Pitman, R. K., Orr, S. P., Altman, B., Longpre, R. E., Poiré, R. E., Macklin, M. L., Michaels, M. J. and Steketee, G. S. (1996). Emotional processing and outcome of imaginal flooding therapy in Vietnam veterans with chronic posttraumatic stress disorder, *Compr. Psychiat.* **37**(6), 409–418.

Pourtois, G., De Gelder, B., Bol, A. and Crommelinck, M. (2005). Perception of facial expressions and voices and of their combination in the human brain, *Cortex* **41**, 49–59.

Pourtois, G., De Gelder, B., Vroomen, J., Rossion, B. and Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information, *Neuroreport* **11**, 1329–1333.

Riva, G. (2005). Virtual reality in psychotherapy: Review, *Cyberpsychol. Behav.* **8**, 220–230; Discussion 231–240. DOI:10.1089/cpb.2005.8.220

Robins, D. L., Hunyadi, E. and Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues, *Brain Cognition* **69**, 269–278. DOI:10.1016/j.bandc.2008.08.007

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S. and North, M. (1995). Virtual reality graded exposure in the treatment of acrophobia: A case report, *Behav. Ther.* **26**, 547–554.

Rudrauf, D., Lachaux, J.-P., Damasio, A., Baillet, S., Hugueville, L., Martinerie, J., Damasio, H. and Renault, B. (2009). Enter feelings: Somatosensory responses following early stages of visual induction of emotion, *Int. J. Psychophysiol.* **72**, 13–23. DOI:10.1016/j.ijpsycho.2008.03.015

Sarlat, L., Warusfel, O. and Viaud-Delmon, I. (2006). Ventriloquism aftereffects occur in the rear hemisphere, *Neurosci. Lett.* **404**, 324–329. DOI:10.1016/j.neulet.2006.06.007

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

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

Stein, B. E. and Meredith, M. A. (1993). *The Merging of the Senses*. MIT Press, Cambridge, MA, USA.

Stein, B. E. and Stanford, T. R. (2008). Multisensory integration: Current issues from the perspective of the single neuron, *Nat. Rev. Neurosci.* **9**, 255–266. DOI:10.1038/nrn2331

Streiner, D. L. and Norman, G. R. (2011). Correction for multiple testing: Is there a resolution? *Chest* **140**, 16–18. DOI:10.1378/chest.11-0523

Suied, C., Drettakis, G., Warusfeld, O. and Viaud-Delmon, I. (2013). Auditory–visual virtual reality as a diagnostic and therapeutic tool for cynophobia, *Cyberpsychol. Behav. Soc. Netw.* **16**, 145–152. DOI:10.1089/cyber.2012.1568

Tanaka, A., Koizumi, A., Imai, H., Hiramatsu, S., Hiramoto, E. and De Gelder, B. (2010). I feel your voice. Cultural differences in the multisensory perception of emotion, *Psychol. Sci.* **21**, 1259–1262. DOI:10.1177/0956797610380698

Thyer, B. A., Papsdorf, J. D., Davis, R. and Vallecorsa, S. (1984). Autonomic correlates of the subjective anxiety scale, *J. Behav. Ther. Exp. Psy.* **15**, 3–7.

Vagnoni, E., Lourenco, S. F. and Longo, M. R. (2012). Threat modulates perception of looming visual stimuli, *Curr. Biol.* **22**, R826–R827. DOI:10.1016/j.cub.2012.07.053

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

Viaud-Delmon, I., Venault, P. and Chapouthier, G. (2011). Behavioral models for anxiety and multisensory integration in animals and humans, *Progr. Neuro-Psychopharmacol. Biol. Psychiat.* **35**, 1391–1399. DOI:10.1016/j.pnpbp.2010.09.016

Viaud-Delmon, I., Znaïdi, F., Bonneel, N., Doukhan, D., Suied, C., Warusfel, O., Nguyen, K. V. and Drettakis, G. (2008). Auditory–visual virtual environments to treat dog phobia, in: *Proceedings of the 7th ICDVRAT — International Conference on Disability, Virtual Reality and Associated Technologies*, Maia, Portugal, pp. 119–124. University of Reading, Reading, UK.

Vines, B. W., Krumhansl, C. L., Wanderley, M. M., Dalca, I. M. and Levitin, D. J. (2011). Music to my eyes: cross-modal interactions in the perception of emotions in musical performance, *Cognition* **118**, 157–170. DOI:10.1016/j.cognition.2010.11.010

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

Vroomen, J., Driver, J. and De Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cogn. Affect. Behav. Neurosci.* **1**, 382–387.

Wald, J. and Taylor, S. (2001). Efficacy of virtual reality exposure therapy to treat driving phobia: A case report, *J. Behav. Ther. Exp. Psychiat.* **31**, 249–257.

Wolpe, J. (1973). *The Practice of Behavior Therapy*, 2nd edn. Pergamon, New York.