Université Pierre et Marie Curie

Ecole Doctorale n°158 : Cerveau, Cognition, Comportement

Laboratoire Sciences et Technologies de la Musique et du Son UMR 9912

Equipe Espaces Acoustiques et Cognitifs

Inducing feelings of fear with virtual reality

The influence of multisensory stimulation on negative emotional experience

Par **Marine TAFFOU**

Thèse de Doctorat de Neurosciences

Dirigée par Isabelle VIAUD-DELMON

Présentée et soutenue publiquement le 17 décembre 2014

Devant un jury composé de :

Olaf BLANKE	Professeur des Universités - Praticien Hospitalier, EPFL, Suisse	Rapporteur
Stéphane BOUCHARD	Titulaire de la Chaire de Recherche du Canada en Cyberpsychologie Clinique Professeur régulier à l'UQO, Canada	Rapporteur
Yann COELLO	Professeur des Universités - Université Lille III, France	Examinateur
Stéphanie DUBAL	Chargé de Recherche, CNRS, France	Examinateur
Vincent HAYWARD	Professeur des Universités, UPMC, France	Examinateur
Ana TAJADURA-JIMÉNEZ	ESRC Future Research Leader UCL, United Kingdom	Examinateur
Isabelle VIAUD-DELMON	Directeur de Recherche, CNRS, France	Directeur

Equipe Espaces Acoustiques et Cognitifs
Laboratoire Sciences et Technologies de la Musique et du Son (STMS)
UMR 9912 IRCAM CNRS UPMC
Institut de Recherche et Coordination Acoustique/Musique
1 place Igor Stravinsky, 75004 Paris - France

Ecole doctorale n°158 – ED3C
7 quai Saint Bernard
Bat. B, 1er étage, porte 115, case 32
75005 Paris - France

I dedicate this thesis to my family

Acknowledgements

I would like to express the deepest appreciation to all those who made this thesis possible and an amazing experience for me.

First, I would like to warmly thank my supervisor, Isabelle Viaud-Delmon, for her advices and encouragements throughout these three years. Thank you for your support and for all you have taught and shared with me. Working with you was a very stimulating and very pleasant experience!

I am very grateful to all the members of my thesis committee for agreeing to evaluate this work. I thank you for your interest and your time and I hope you will enjoy reading the manuscript.

I would like to thank Hugues Vinet and Gérard Assayag for having welcomed me into the *Music and Sound Sciences and Technologies* lab at l'Institut de Recherche et Coordination Acoustique/Musique (IRCAM). I felt very privileged to benefit from the amazing work environment of the lab and of the *Acoustic and Cognitive Spaces* team. Many thanks to Olivier Warusfel and to all the people of the *Acoustic and Cognitive Spaces* team, for their kindness and availability. I am delighted to be part of the team. Thank you also for your major contribution to the auditory virtual environments and your precious help in all technical matters. I also have a special thought for the two undergraduate students who contributed to this research. Thank you for your hard work and your energy!

I also had the chance to collaborate with the *Social and Affective Neuroscience* (SAN) team lead by Nathalie George and Philippe Fossati at the Centre de Recherche de l'Institut du Cerveau et de la Moëlle épinière (CRICM). I am very grateful to all the members of the SAN team for their warm welcome. I have greatly benefited from their insightful comments and suggestions on my work and from the pleasant and stimulating environment of the team. I would particularly like to thank Stéphanie Dubal who transmitted me her enthusiasm for emotion research during my Master internship and who introduced me to Isabelle. This was the starting point of this great experience, which would not have been possible without you and Isabelle. Thank you very much for your precious help and advice over the years.

The present research was part of a European project, which aimed to develop technologic tools to support the treatment of people who are at risk of social exclusion due to fear and/or apathy associated with a disability (VERVE project coordinated by Trinity College Dublin, 2011-2014, http://www.verveconsortium.eu/). I thank all of the collaborators of the project and particularly the *Rendering and Virtual Environments with Sound* (REVES) team from l'Institut national de recherche en informatique et en automatique (INRIA) and the *Graphics Vision Visualisation* (GV2) team from Trinity College Dublin (TCD) for their work in the development of the virtual environments I used in my work and for helping me set up my experiments in virtual reality.

I would like to thank the following institutions, organizations and their staff: l'Institut de Recherche et Coordination Acoustique/Musique (IRCAM) and l'Institut du Cerveau et de la Moëlle épinière (ICM) who hosted me during my thesis; and the European Commission who funded the research. The present research work was supported by the EU FP7-ICT-2011-7 project VERVE, grant n°288910.

I also thank all the persons who took part in my experiments, for their interest, patience and good will.

Last, but not least, I warmly thank all the members of my family and all my friends for their unconditional support. Special thanks to Florent, who was always there to cheer me up!

Table of Contents

Table of Figures

Table of Tables

INTRODUCTION

The present research work aimed at further understanding human affect. It was inspired by the field of clinical research on virtual-reality based therapy for phobias. This field of research investigates the effectiveness of presenting virtually the fear-object to patients in order to reduce their fear with exposure therapy. Findings from this field revealed that exposing patients to the object of their fear gradually in virtual reality is effective in treating different phobias. The gradation of the intensity of exposure is implemented by first presenting the object of fear in conditions where it induces relatively low feelings of fear in the patient. Then progressively, the conditions of the fear object presentation are modified so as to attain the condition that induces the most feeling of fear in the patient. Different conditions where the participants fear the object more or less often differ in terms of context, spatial and/or sensory characteristics. Virtual reality allows complete control over the presentation characteristics of the fear objects, thus representing an advantageous media for exposure therapy. While the characteristics of fear-objects presentation seem to play a role in the intensity of subjects' fear, empirical studies investigating the impact of stimulus presentation on conscious emotional experience remained sparse. A further understanding of how the characteristics of presentation of the fear-objects modulate the subjects' feelings of fear could help refine the design of virtual environments for exposure therapy and further exploit the advantages of virtual reality for the treatment of phobias.

In this work, I conducted three studies to investigate how the sensory presentation of feared objects influences feelings in virtual reality. The findings contribute to the field of human affect research and to the field of virtual-reality based therapy for phobias research.

This manuscript is divided in two parts: the first part introduces the theoretical framework and the second part describes the three studies that were conducted and discusses them separately and generally. Two of these studies have been published; the corresponding papers are inserted in chapter 6 and 7, in the second part of the manuscript. The third study is still ongoing work and is detailed in chapter 8.

1. VIRTUAL REALITY

For the past two decades, the interest in virtual reality as a tool for therapy and research has grown. Pioneer research work using virtual reality techniques emerged in the field of clinical psychology (Hodges et al., 1995; Hodges, Watson, Kessler, Rothbaum, & Opdyke, 1996; North, North, & Coble, 1997, 1998; Rothbaum et al., 1995a). These pioneer studies explored the effectiveness of the use of computer-generated, virtual environments for the therapy of phobias. The first phobias, which were targeted, were acrophobia (the fear of heights) and flight phobia (the fear of flying).

A standard treatment for phobias is exposure therapy, which intends to reduce fear responses and experiences in feared situations. It consists of a progressive confrontation with fearful situations along several therapeutic sessions with the objective of triggering a habituation phenomenon. Traditionally, this exposure is conducted *in vivo* with patients facing real feared situations. In the aforementioned studies, the exposure was conducted with feared situations presented virtually to the subjects within virtual environments (*in virtuo*). The success of the exposure in virtual environments in reducing fear of heights and fear of flying put virtual reality forward as a new medium for the treatment of phobias. From these studies, the interest for virtual reality as a therapeutic tool has emerged.

One major advantage that virtual reality offers for exposure therapy is the ability to completely control the feared situation or object presented to the subject. Different physical parameters of the virtual stimulation can be manipulated. For example, the amount of sensory information delivered to the subjects and/or the location of the feared stimuli in the virtual space and in relation to the subject can be controlled and manipulated. The manipulation of these sensory and spatial parameters has intuitively been used in the design of virtual environments for the treatment of phobias in order to modulate the intensity of exposure, i.e. the intensity of fear induced in the subjects. For instance, in the virtual environment for flight phobia, the sound of the activation of the airplane engine is added to the virtual stimulation to increase the intensity of exposure. In the virtual environment for acrophobia, the psychologist increases the intensity of exposure by placing the subjects closer to the edge of a glass elevator.

A further understanding of how the sensory and spatial characteristics of virtual stimulation influence the intensity of fear induced in subjects could help further exploiting the advantage of virtual reality when designing virtual environments and scenarii for the treatment of phobias. Moreover, virtual reality seems to be a particularly appropriate tool to investigate the links between multisensory stimulation and conscious emotional experience in space.

Virtual reality techniques

The term **virtual reality** (VR) refers to a set of technologies, which allow for the immersion of individuals in computer-simulated environments. With virtual reality techniques, individuals can be placed in three dimensional, complex, dynamic and interactive virtual environments (VE) depicting imagined or real places. VR involves techniques that engender real-time rendering of sensory information from different modalities (visual, auditory, haptic, proprioceptive) and tracking systems enabling appropriate sensory rendering with respect to the user's movements. This interactive sensory rendering aims at inducing a feeling of presence in the VE in the user (Schubert, Friedmann, & Regenbrecht, 2001). Different display systems can be used to achieve this goal (see Figure 1.1 for examples).

Figure 1.1. Examples of different display systems to render visual information.
Virtual visual information can be presented in 3D through a Head-Mounted Display (HMD; left, retrieved from Rothbaum et al.,1995), on a stereoscopic passive screen with the user wearing 3D glasses (middle), or on a four-sided retro-projected cube (right) with the users also wearing 3D glasses.

Virtual reality and fear

Since 1995, the investigation of the utility of VR for emotional rehabilitation of phobias has grown (see Cote & Bouchard, 2008 for a review). The effect of exposure therapy *in virtuo* on treatment outcome was evaluated for different phobias with the measures typically used in traditional exposure therapy *in vivo*. These measures generally involve questionnaires evaluating the subjects' fear and behavioral avoidance tests (BAT, also called behavioral assessment test). The BAT measures the behavioral component of subjects' fear. During the BAT, the subject is confronted with the fear object and is asked to complete a series of tasks, which are progressively more anxiogenic. Generally, this involves getting progressively closer to the fear-object as, for example, one of the first descriptions of a BAT measuring the fear of snakes can attest (Lang & Lazovik, 1963). The level of fear is evaluated by the number of tasks that the subject is able to undergo or how close he/she is able to approach the fear-object. This score can be compared before and after therapy to assess its success. Collecting subjective reports of experienced fear at each stage of the BAT, can also be used to assess the level of fear. The typically-used report is the Subjective Units of Distress (SUD: Wolpe, 1973). SUD is a self-report measurement of experienced fear or discomfort, which has been shown to correlate with physiological measures of arousal state (Thyer, Papsdorf, Davis, & Vallecorsa, 1984). SUD measure rates the level of experienced fear or discomfort on a scale from 0 to 100. Studies on exposure therapy in VR have either conducted this BAT *in vivo* (e.g. Rothbaum et al., 1995) or *in virtuo* (e.g. Mühlberger, Sperber, Wieser, & Pauli, 2008).

Successful outcome of exposure therapy in VR has been found, using different VEs (see examples Figure 1.2), for several specific phobias, including arachnophobia (e.g. Carlin, Hoffman, & Weghorst, 1997; Garcia-Palacios, Hoffman, Carlin, Furness, & Botella, 2002), cockroach phobia (e.g. Botella, Bretón-López, Quero, Baños, & García-Palacios, 2010), acrophobia (e.g. Choi, Jang, Ku, Shin, & Kim, 2001; Coelho, Santos, Silvério, & Silva, 2006; Emmelkamp et al., 2002; Emmelkamp, Bruynzeel, Drost, & Van der Mast, 2001; Krijn et al., 2004; Rothbaum et al., 1995b), claustrophobia (e.g. Botella et al., 1998; Botella, Baños, Villa, Perpiñá, & García-Palacios, 2000; Botella, Villa, Banos, Perpina, & Garcia-palacios, 1999), fear of flying (e.g. Mühlberger, Herrmann, Wiedemann, Ellgring, & Pauli, 2001; Rothbaum et al., 2006; Wiederhold, Gevirtz, & Wiederhold, 1998) and fear of driving (e.g. Wald & Taylor, 2001), and also for other anxiety disorders such as social phobia (e.g. Anderson, Rothbaum, & Hodges, 2003), post-traumatic stress disorder (e.g. Beck, Palyo, Winer, Schwagler, & Ang, 2007) and panic disorder (e.g. Botella et al., 2007). The efficacy of exposure therapy in VR is assumed to be linked to three factors: (1) the possibility of navigation in the VE, (2) the induction of affective states by the virtual feared stimuli and (3) the fact that modifications in behaviors and feelings can be generalized to real situations.

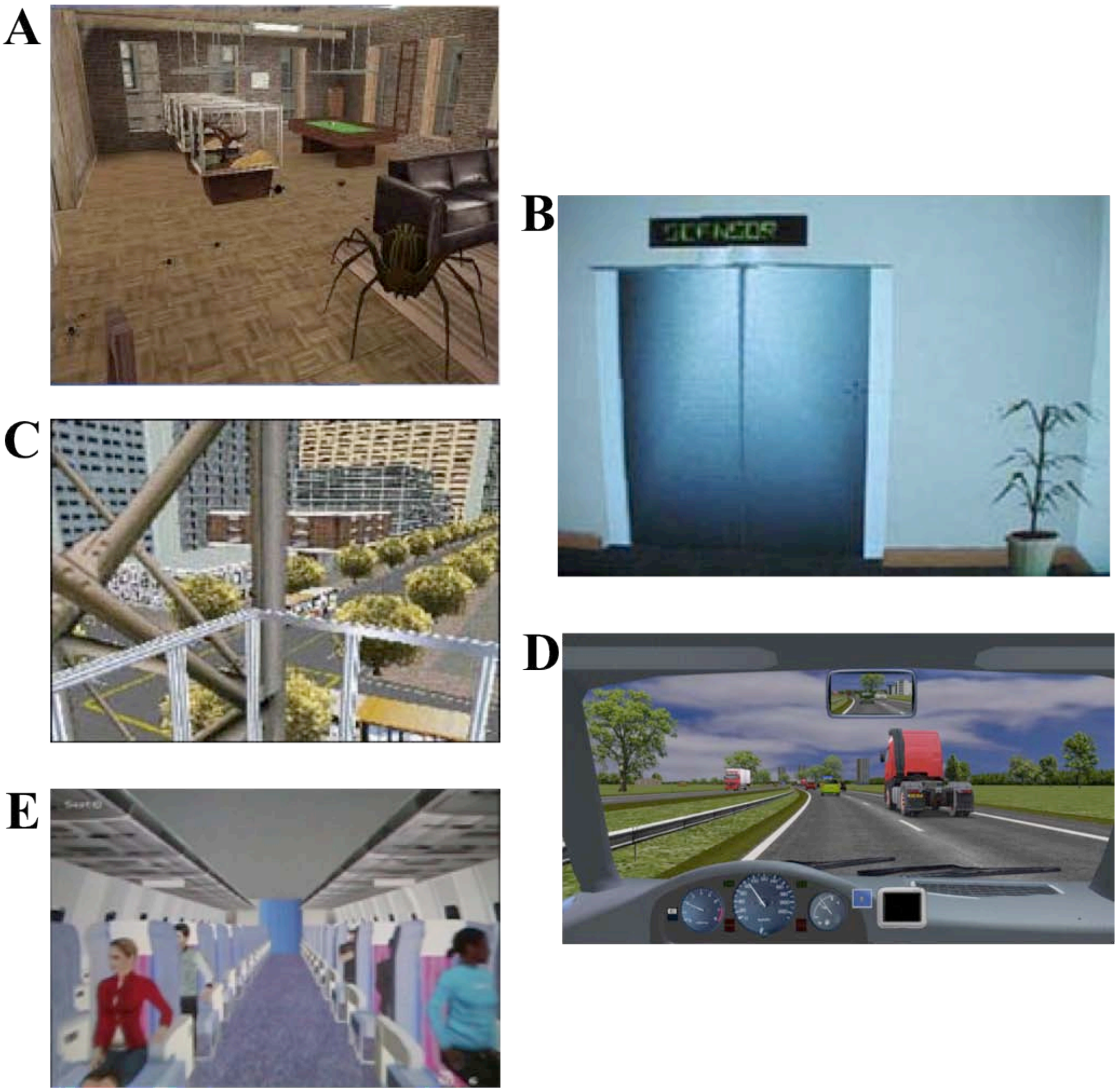

Figure 1.2. Examples of virtual environments used to address different phobias
Virtual environments to address arachnophobia (A), claustrophobia (B), acrophobia (C), fear of driving (D) and fear of flying (E).
The pictures have been retrieved from http://www.vrphobia.com/therapy.htm and http://www.stsoftware.nl/rijangst.html

The enthusiasm for the use of VR in emotional rehabilitation is due to the numerous advantages that it provides for the treatment of anxious disorders (North et al., 1998). VR allows for the exposure of patients to feared stimuli, which are complex, dynamic, interactive and in 3D. The feared stimuli or situations are totally controlled, preventing unpredicted events from interfering with treatment. Situations can also be repeated and the intensity of exposure manipulated, enabling the establishment of a treatment plan and its enaction in total safety for the patient. Additionally, the privacy and confidentiality of treatment is preserved given that patients and therapist remain in the therapy office for exposure in VR. Furthermore, the attractiveness of VR increases the propensity to seek treatment and decreases drop offs, leading to higher probability of therapeutic success. The use of VR has also been extended to the research in the treatment of other psychiatric disorders (e.g. Riva, Bacchetta, Baruffi, Rinaldi, & Molinari, 1999 for anorexia; Saladin, Brady, Graap, & Rothbaum, 2006 for substance dependence) and in cognitive rehabilitation (e.g. Kim, Chun, Yun, Song, & Young, 2011).

Virtual reality, multisensory integration and spatial behaviors

More recently, the interest for VR has started to grow in neuroscience research (see Bohil, Alicea, & Biocca, 2011). One of the challenges in neuroscience research is to design experimental paradigms allowing for the examination of behaviors and underlying cerebral processes in natural situations while enabling experimental control. Highly controlled experimental design provides an efficient strategy to disentangle different processes by precisely submitting different variables for study. However, this high experimental control implies a simplification of natural stimuli, which cannot totally account for stimulation coming from the real world. On the other hand, paradigms with high ecological validity take the processes in real situations into consideration, but allow for only very weak experimental control. VR techniques provide a tool that allows for the design of middle-ground paradigms. VR simulates naturalistic environments, in which stimuli are embedded in a meaningful context. In VR paradigms, many variables of stimulation can be manipulated and controlled: the timing, the complexity of stimuli and also the dynamic interaction of stimuli with the user.

Specifically, VR is in essence a multisensory tool and easily allows the manipulation of different sensory inputs delivered to the user, thus providing an ideal tool for research in multisensory integration. This advantage is already being used for the investigation of multisensory perception of external objects (e.g. Suied, Bonneel, & Viaud-Delmon, 2009) as well as for exploring multisensory integration in bodily self consciousness (e.g. Lenggenhager, Tadi, Metzinger, & Blanke, 2007; Slater, Spanlang, Sanchez-Vives, & Blanke, 2010). In the field of clinical psychology and emotional rehabilitation, the multisensory capacity of VR has often been underexploited. For example, in the VEs, the auditory modality is often absent or is not rendered in interactive 3D and only used to deliver simple associative cues with the complex visual stimulation rendered in 3D.

VR has also been used in the investigation of spatial behaviors. The interactive quality of VR makes navigation in a three-dimensional virtual space possible. This spatial capacity has already been exploited by several studies on spatial cognition (e.g. Driscoll, Hamilton, Yeo, Brooks, & Sutherland, 2005) and spatial behavior during social interaction (e.g. Jeffrey & Mark, 1998; Wilcox, Allison, Elfassy, & Grelik, 2006). Moreover, the studies investigating emotional rehabilitation of phobias often take advantage of the spatial capacity of VR when controlling the graduated intensity of exposure by manipulating the distance between the patient and the fear object.

Challenges of virtual reality

Although VR has many advantages for research and therapy, some difficulties also come along with its use. Besides the technological complexity, often requiring specialist technology skill, and the cost of the rendering systems, VR can induce transient unpleasant side effects. During navigation within VE, the user can experienced a sort of motion sickness with dizziness, nausea or headache. This motion sickness in VR is referred to as *cybersickness* and is certainly related to sensory conflicts during navigation in VE; for example, whereas the visual information indicates movement, proprioceptive information indicates stasis (Bos, 2007).

Over the past 20 years, virtual reality has emerged as an interesting tool for the treatment of phobias. Within virtual environments, the spatial locations and sensory presentation of feared stimuli can be controlled and manipulated in order to modify the intensity of exposure during treatment. Moreover, virtual reality represents a perfect tool for the investigation of sensory and spatial determinants of emotional experience. The use of virtual environments in order to display naturalistic stimuli embedded within a significant context allows for new empirical approaches at the intersection of ecological validity and experimental control.

Investigating the influence of spatial and sensory parameters of feared stimuli on the conscious emotional experience they induce in the subjects with virtual reality could help further exploit the advantages of virtual reality for the treatment of phobias.

2. MULTISENSORY PROCESSING

2.1. Multisensory integration

We perceive the world *via* multiple senses. When a dog is happy to see you, you can see him jump towards you, feel his paws on your thighs, hear him panting from excitement and even smell his breath. Even though these cues about the dog's presence are delivered to different senses, we perceive the dog as a singular object of the external world. The sensory information coming from vision, touch, audition and smell are combined, integrated into a unique percept.

Although we almost constantly integrate multisensory information, we tend to not be aware that this phenomenon occurs. However, there are particular situations, in which we can witness multisensory integration. In these situations, the discrepancies between the information coming from different senses help reveal the multisensory integration processes at stake. The most famous example is ventriloquism. In this situation, the ventriloquist generates a speech sound without lip movement whilst moving the lips of a puppet that he holds close to him, in accordance with the speech he produces. When the members of the audience perceive the performance, they have the illusion that the puppet is speaking (Figure 2.1). The perceived spatial localization of the speech sound is shifted toward the location of the lip movements corresponding to the production of the sound i.e. the puppet's lip movements. This bias provides evidence for multisensory integration (Calvert, Spence, & Stein, 2004). The illusion exposes the interaction between the cues from the visual and auditory modalities, which occurs when perceivers interpret the spatial location of the speaker. Research has profited from cross-modal biases, such as the bias observed in the ventriloquism illusion, for the study of multisensory integration.

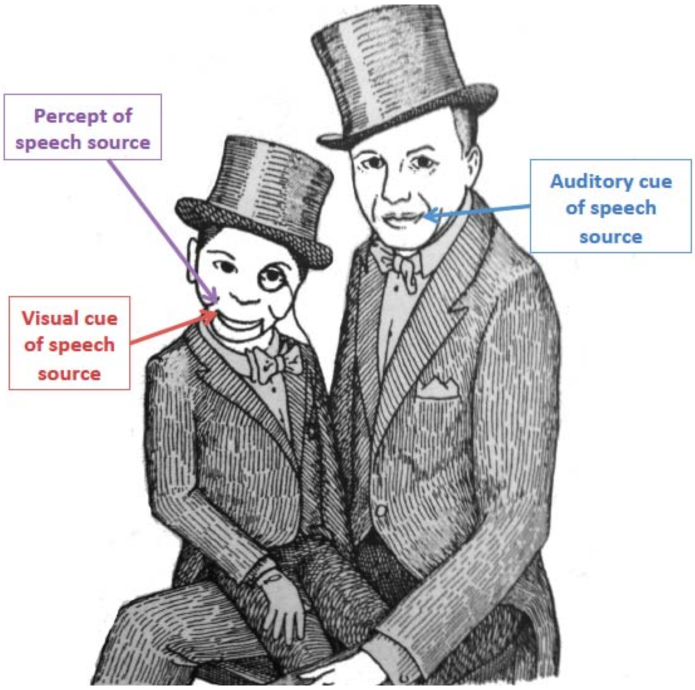

Figure 2.1. The ventriloquism effect (adapted from Stein & Meredith, 1993).
The ventriloquist produces a speech sound without moving his lips whilst moving the lips of the puppet. The cues coming from audition and vision deliver discordant information concerning the location of the speech source. The auditory cue indicates the ventriloquist as the speaker (red arrow) while the visual cue indicates the puppet as the speaker (blue arrow). The audience perceives the speech as coming from the puppet (purple arrow). This illusion reveals an interaction between auditory and visual information, which occur when the audience interprets the spatial location of the speaker.

In the multisensory research field, the ventriloquism effect (Howard & Templeton, 1966) has been studied using paradigms consisting of presenting a visual and an auditory stimulus at the same time but at slightly different spatial locations and assessing participants' perceived location of the event that the two stimuli constitute. Using these paradigms, the integration of auditory and visual cues has not only been demonstrated with meaningful complex stimuli (e.g. Pick, Warren, & Hay, 1969) but also with simpler stimuli such as beep sounds and flashes of light (e.g. Bertelson & Radeau, 1981).

Automatic multisensory integration

Auditory-visual integration seems to be automatic. The spatial cross-modal bias is not linked to the observers' decision to integrate the spatial cues from the visual and auditory modalities because they are aware of the discordances between them (Bertelson & Aschersleben, 1998).

Moreover, a shift of the perceived sound location toward the visual stimulus location is still observed even when participants are instructed to only focus on localizing the auditory stimulus while trying to ignore the stimulus from the visual modality (Bertelson & Radeau, 1981), suggesting that the direction of attention does not influence auditory-visual integration. Further studies have supported this assumption by demonstrating that the ventriloquism effect is independent from the direction of attention: be this deliberate (Bertelson, Vroomen, De Gelder, & Driver, 2000) or automatic (Vroomen, Bertelson, & De Gelder, 2001).

What is more, the automaticity of sensory integration is not limited to auditory-visual events. Studies using different cross-modal bias paradigms have reported the same conclusion for visuo-tactile (Bresciani, Dammeier, & Ernst, 2006) and visuo-haptic integration (Helbig & Ernst, 2008).

Spatial, temporal and semantic determinants of multisensory integration

Even if the multisensory integration seems automatic, this does not mean that we systematically integrate every pair of sensory inputs that we perceive. Multisensory integration often requires that the sensory cues occur close in time and in space. The ventriloquism illusion would certainly disappear if the performer moved the puppet's lips only after he has finished speaking or if the puppet had been located at 2m from him.

The range of temporal and spatial distance between cues, which allow for the perception of a unique percept, has been approximated for neutral stimuli. In a study wherein participants had to judge the likelihood that a sound burst and a flashing light spot – both presented with different spatial and temporal disparities – have a common cause, the converging point of subjective spatial alignment was located at positions where visual and auditory stimuli are in exact objective spatial alignment (Lewald & Guski, 2003). Concerning the point of subjective simultaneity, subjects found that the synchrony is at an optimal level when the visual stimulus is presented 90ms before the auditory stimulus. This phenomenon is partly due to the fact that the transduction of visual signals in the retina is slower than auditory transduction processes (Fain, 2003). A spatial disparity from -7° to +7° and a temporal disparity, with the visual stimulus preceding the auditory stimulus, from -25ms to 205ms are tolerated between sensory cues. Outside this spatio-temporal window, the two sensory inputs are judged as caused by different events. These results have been obtained with simple meaningless stimuli. The spatio-temporal window may be wider with complex stimuli with semantic content.

Moreover, the semantic congruency also seems to constrain multisensory integration (Laurienti, Kraft, Maldjian, Burdette, & Wallace, 2004; Miller, 1991). Here again, we can perfectly imagine that the ventriloquism illusion would not occur if the performer moved the feet instead of the lips of the puppet.

* * *

When perceiving multisensory events in the external world, the different sensory cues they deliver are automatically combined into a unified percept. Spatial, temporal and semantic factors are taken into account when connecting the different sensory cues in order to create meaningful combinations.

2.2. Neural consequences of multisensory stimulation

For the past 20 years, the research activity on multisensory integration has increased (see Alais, Newell, & Mamassian, 2010, for a broad review of the field). In particular, the study of the neural correlates of multisensory processing has significantly grown since Stein and colleagues started to lead the march with a body of work studying the mammalian superior colliculus response to multisensory stimuli (Stein & Meredith, 1993). This section provides an overview of what is known about the influence of multisensory stimuli on cerebral processing and introduces the key principles of multisensory integration, which have been established in the past two decades.

2.2.1. Superior Colliculus responses to multisensory stimuli

For multisensory integration to happen, sensory inputs must converge onto single neurons or ensembles of interconnected neurons. The superior colliculus (SC) is a sub-cortical structure of the brain, which receives visual, auditory and somatosensory inputs (Meredith, Nemitz, & Stein, 1987). The neurons of SC deep layer are responsive to inputs from two, or even three, sensory modalities. The receptive fields of these neurons overlap for the different sensory modalities so that they respond to inputs according to their spatial location. In other words, an auditory, a visual and a somatosensory input located within a same region of space will activate the same neuron. These multisensory neurons of the SC consequently represent a good target for the investigation of how multisensory stimuli integrate.

Stein and colleagues studied the physiological response to multisensory events at the level of single neurons, in the cat SC. They measured the activity of neurons in response to multisensory simple events (e.g. a moving light bar coupled to a hiss sound) with extracellular recording techniques. They observed significantly higher activity in response to multisensory events than in response to unimodal events. For example, the average of impulses evoked in neurons by an auditory-visual event was significantly higher than the one evoked by the corresponding auditory or visual unimodal event. The multisensory responses even exceeded the sum of the unimodal responses, an effect that Stein et al. called "superadditivity" (see Figure 2.2). This **response enhancement** (Stein & Meredith, 1993) was found for every multisensory category (auditory-visual, somatosensory-auditory, somatosensory-visual and trimodal).

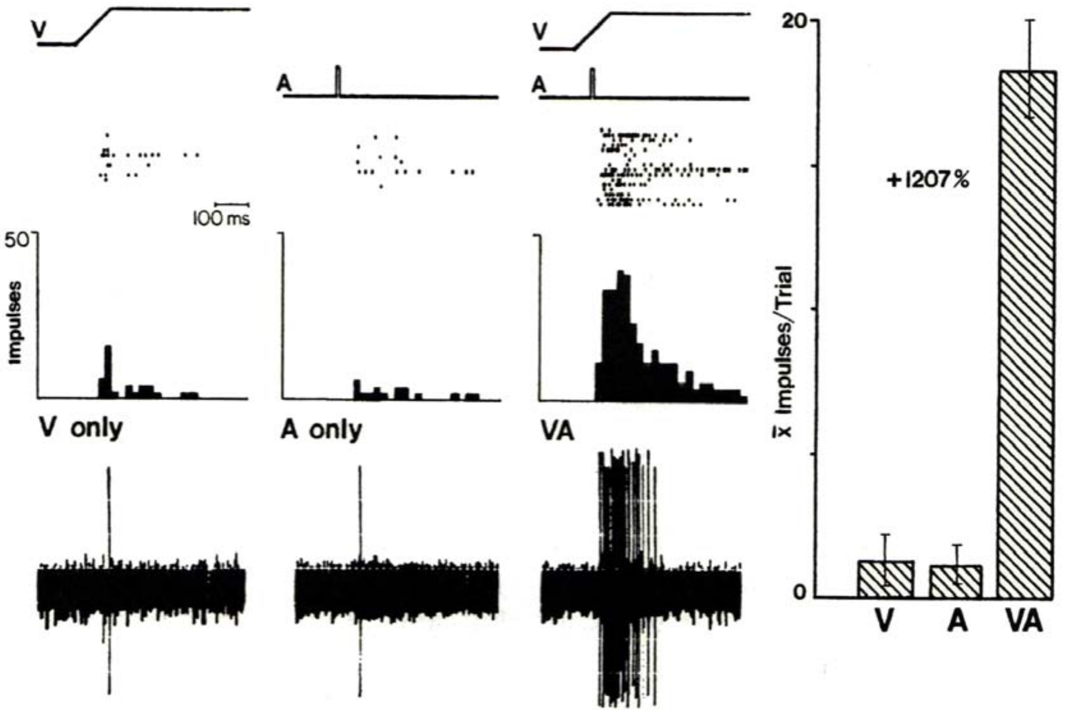

Figure 2.2. Auditory-visual response enhancement in a neuron of the cat superior colliculus (from Stein & Meredithn 1993).
This figure depicts the responses evoked by visual (V), auditory (A) and auditory-visual (VA) stimuli in a neuron of the cat superior colliculus. Responses are displayed in the impulse rasters (in which each dot represents a single neural impulse and each row represents a single trial) as well as in peri-stimulus time histograms (in which the impulses are summed across trials at each moment of time and binned) and single-trace oscillograms below the stimulus traces. While the visual and auditory stimuli evoke weak responses in the neuron, their combination produces strong responses on every trial. The mean number of impulses per trial (right histogram) in response to multisensory stimulation greatly exceeds the one in response to either stimulus alone. This response enhancement is even superadditive because the auditory-visual response exceeds the sum of the visual and auditory responses.

However, if the sensory inputs were spatially discordant (in terms of neurons receptive fields) during multisensory stimulation, the inputs were processed as separate events and the response enhancement did not occur. Instead, an opposite phenomenon often occurs: a response depression. The response enhancement also disappeared if there was a substantial temporal discrepancy between the sensory inputs.

The enhancement of the neurons responses with multisensory events was found to be variable depending on the effectiveness of each unimodal stimulus. An effective unimodal stimulus evokes, on its own, high responses in the neuron while an ineffective unimodal stimulus evokes low or even no responses in the neuron. The increase of the neuron responses with bimodal events coupling two effective unimodal stimuli was small in relation to the responses with the individual stimuli. The response enhancement was subadditive, i.e. the response to bimodal stimuli was higher than the response to each unimodal stimuli but lower than their sum. Contrastingly, bimodal events composed of two ineffective unimodal stimuli evoked substantially higher responses as compared to the modest responses they had induced in the neuron when presented individually. The response enhancement was superadditive, i.e. the response to bimodal stimuli was higher than the response to each unimodal stimulus and even exceeded the sum of the unimodal responses. Furthermore, two stimuli, which were incapable of evoking any responses on their own, may induce a response in the neuron when combined. It seems that the magnitude of the response enhancement with multisensory events increases as the effectiveness of the individual sensory inputs decreases (Figure 2.3). This rule, known as the "**inverse effectiveness principle**" (Stanford, Quessy, & Stein, 2005), suggest that maximal response enhancement occurs when the responsiveness to individual sensory inputs are minimal. This principle illustrates the notion that multisensory enhancement is most useful in situations where none of the individual sensory stimuli are effective enough to guarantee detection.

2.2.2. Cortical responses to multisensory stimuli

With their work, Stein and colleagues established some key principles of multisensory processing and guided multisensory investigation in cortical areas. Cortical areas in which sensory inputs from different modalities converge and whose responses to multisensory stimuli meet the criterion for multisensory integration, i.e. response enhancement, were looked for.

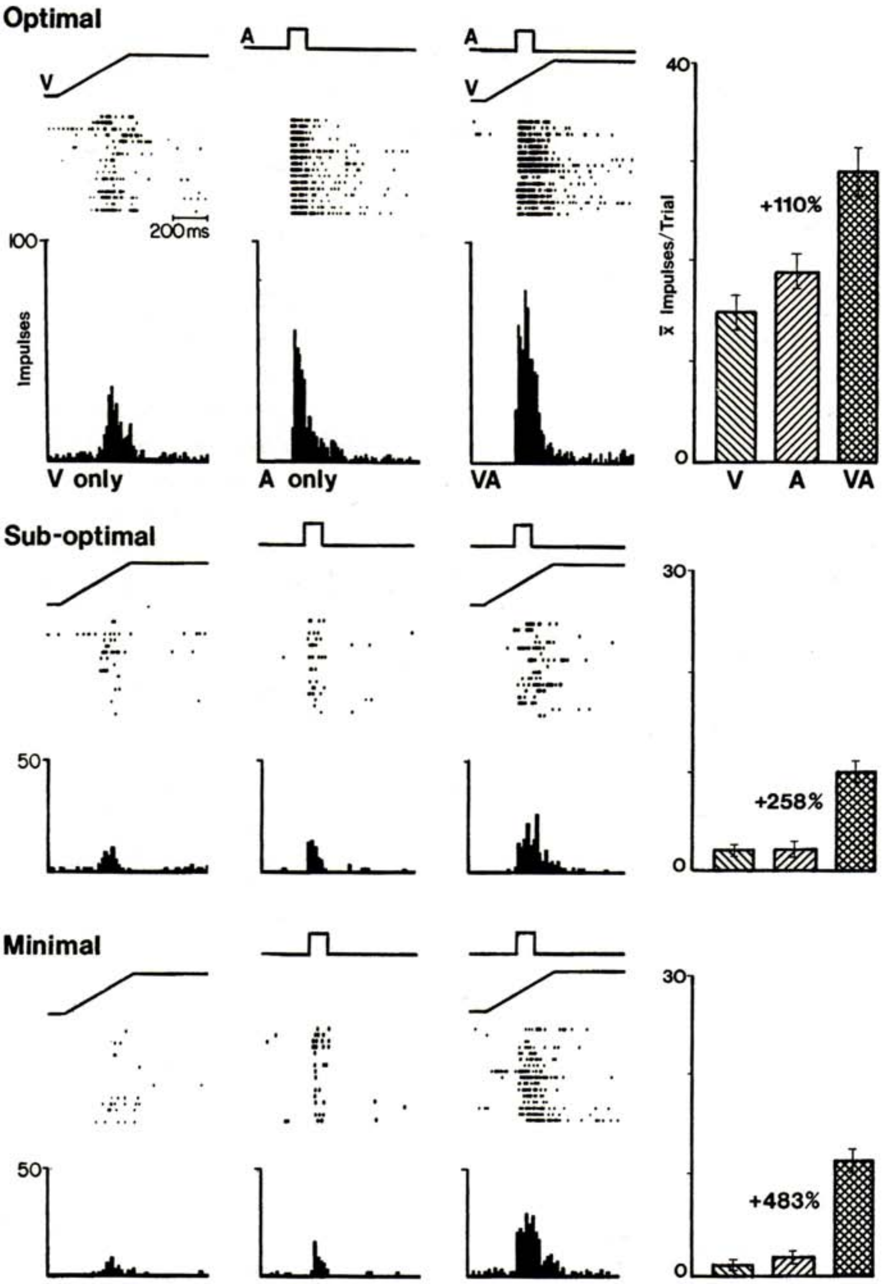

Figure 2.3. Inverse effectiveness of auditory-visual stimulation in a neuron of the cat superior colliculus (from Stein & Meredith, 1993).

The auditory-visual response enhancement increases as unimodal stimulus effectiveness decreases in the neuron. When the effectiveness of the unimodal stimuli is optimal (top), the multisensory response enhancement is subadditive, such that the response exceeds the most effective component response but not their sum. As the unimodal stimuli become less effective, the multisensory enhancement become proportionally higher. The response enhancement becomes additive with sub-optimal unimodal stimuli (middle) and superadditive with minimal unimodal stimuli (bottom). The maximal multisensory response enhancement occurs when the effectiveness of the individual sensory inputs is minimal.

Cortical sites of multisensory integration

Sensory inputs from different modalities converge in different cortical areas such as posterior parietal areas (Duhamel, Colby, & Goldberg, 1998; Rizzolatti, Scandolara, Matelli, & Gentilucci, 1981), superior temporal areas (Beauchamp, Yasar, Frye, & Ro, 2008; Jiang, Lepore, Ptito, & Guillemot, 2004a; Jones & Powell, 1970), the anterior ectosylvian sulcus (Reinoso-Suarez & Roda, 1985), and even primary sensory cortices (Clavagnier, Falchier, & Kennedy, 2004; Falchier, Clavagnier, Barone, & Kennedy, 2002; Rockland & Ojima, 2003). Neuroimaging and neurophysiological studies on both human and non-human mammalians have investigated the link between responses in these cortical areas and multisensory events.

Using single neuron recordings, Jiang and colleagues studied the response of multisensory neurons located in the anterior ectosylvian cortex of adult cats. The response of these neurons was enhanced when the cat was stimulated with spatiotemporally correlated visual (light bars), auditory (white noise) and somatosensory (jet of air) events (Jiang, Lepore, Ptito, & Guillemot, 2004b). A similar multisensory enhancement was also observed in the posterior parietal areas of the macaque monkey (Avillac, Ben Hamed, & Duhamel, 2007). Providing spatial and temporal coincidence of the sensory inputs, visuo-tactile events evoke an enhancement of the response in most of the neurons of the ventral intraparietal area.

In the human brain, functional magnetic resonance imaging (fMRI) has allowed the examination of cerebral activation in response to multisensory stimuli with high spatial resolution. Calvert and colleagues investigated the cerebral activation in response to auditory-visual speech stimuli. They found a superadditive response to auditory-visual speech in the superior temporal sulcus (STS) when compared to the responses to individual sensory stimulations (Calvert, Campbell, & Brammer, 2000). This multisensory response enhancement was dependent on temporal coincidence: no response enhancement was observed for asynchronous auditory-visual speech. Beauchamp and colleagues also found evidence of auditory-visual integration in the STS. They found an enhancement in STS activation in response to auditory-visual objects, such as animals and tools (Beauchamp, Lee, Argall, & Martin, 2004). Response enhancements in the superior temporal areas have also been observed in response to auditory-tactile events. It has been demonstrated that an auditory noise and a tactile stimulation (delivered by means of a wooden roller device) evoke a superadditive response in the left superior temporal gyrus when presented simultaneously (Foxe et al., 2002).

Multisensory responses in human superior temporal areas also seem to follow the inverse effectiveness principle. By manipulating the effectiveness of auditory-visual stimuli with noise, Stevenson & James demonstrated an inverse correlation between unimodal stimulus effectiveness and multisensory response enhancement in the STS. They found this effect with auditory-visual speech stimuli as well as with auditory-visual videos stimuli of tool use (Stevenson & James, 2009).

Time-course of multisensory integration

Taking advantage of the event related potential (ERP) technique, which provides insights into the time-course of cerebral processes, Giard & Peronnet (1999) researched the processing stages, at which multisensory interactions take place in the human brain. They found that a simple auditory-visual event (an ellipse coupled to a tone burst) modulates cerebral processing as early as 40ms post-stimulation (Giard & Peronnet, 1999). In comparison to the presentation of the ellipse alone or of the tone burst alone, the multisensory event evoked a new neural activity over the right fronto-temporal area from 140ms to 165ms after the presentation of the auditory-visual stimulus. Moreover, an enhanced activity was also observed in the primary sensory cortex areas. If participants processed more effectively the visual rather than the auditory stimulus, they only found an enhancement of the ERP component reflecting the auditory cortex activity (90 to 110ms post-stimulation). If the participant's effectiveness in processing the stimuli was the other way around, they only observed an enhancement of the ERP components reflecting the visual cortex activity (40ms to 90ms post-stimulation). These findings suggest that multisensory integration adaptively induced an enhancement in the sensory processing of the less efficient cue. This is consistent with recent evidence that auditory-visual stimuli response in early cerebral processing follows the inverse effectiveness principle (Senkowski, Saint-Amour, Höfle, & Foxe, 2011).

The modulation of cerebral processes from early stages was also observed in response to auditory-tactile stimuli (Murray et al., 2005). Murray and colleagues found an enhanced cerebral response to auditory-tactile events as early as 50ms post-stimulus with a source localized in the auditory association areas.

* * *

Recently, the prevailing view of multisensory integration occurring at late stages of processing, after the individual processing of each sensory input (Treisman & Gelade, 1980), has been questioned. As reflected by the above overview, the signals coming from the different sensory modalities seem to interact and enhance cerebral activity from very early stages of processing, at multiple stages of processing and in diverse cortical and subcortical brain areas.

2.3. Behavioral consequences of multisensory stimulation

Given that multisensory stimuli enhance cerebral responses, one may wonder whether they influence behavioral responses. Do we take advantage of multisensory information when interacting with the external world? Evidences from behavioral studies suggest that we indeed do. Our performances in several different tasks are better if we are provided with multisensory information.

Faster behavioral responses

Multisensory stimulation leads to faster behavioral responses. For example, we will faster detect and recognize a phone if we can both see it and hear it ring (Suied et al., 2009). This phenomenon is known as the **redundant signal effect** (RSE, Kinchla, 1974): the combined effect of auditory and visual information about the same object leads to shorter reaction times than auditory or visual information alone.

Two main explanations have been proposed to account for this RSE. The increase of detection speediness can be linked to multisensory integration or can be simply related to the fact that more information is available. In the latter case, the inputs coming from different sensory modalities do not need to converge in order to induce a RSE. Raab proposed a model, the **race model** (Raab, 1962), describing how redundant sensory signals would lead to faster detection *via* a statistical facilitation, without converging. Let us again consider the example of the phone. When only hearing the ring tone, the auditory signal processing will accumulate evidence that it is a phone sound until reaching a threshold leading to the recognition of a phone. A similar accumulation of evidences takes place when only seeing the phone. The race model states that the brain never combines evidences from different sensory modalities in order to meet its threshold for detecting or recognizing the object. Instead, the reaction time for the detection or recognition of the auditory-visual phone would be controlled by the sensory signal, which leads to a faster recognition of the phone on its own. Thus, there would effectively be a race between the separate processing of the individual sensory signals from multisensory redundant stimuli (see Figure 2.4A). As the processing time of sensory signals can vary, the reaction time to the winning signal would be, on average, faster than the average reaction time for either racer signal alone. Moreover, the reaction time in response to a multisensory redundant event could be predicted on the basis of the distributions of the reaction times in response to each of the unimodal signal. The fastest possible responses cannot be faster than the fastest possible responses to single signals.

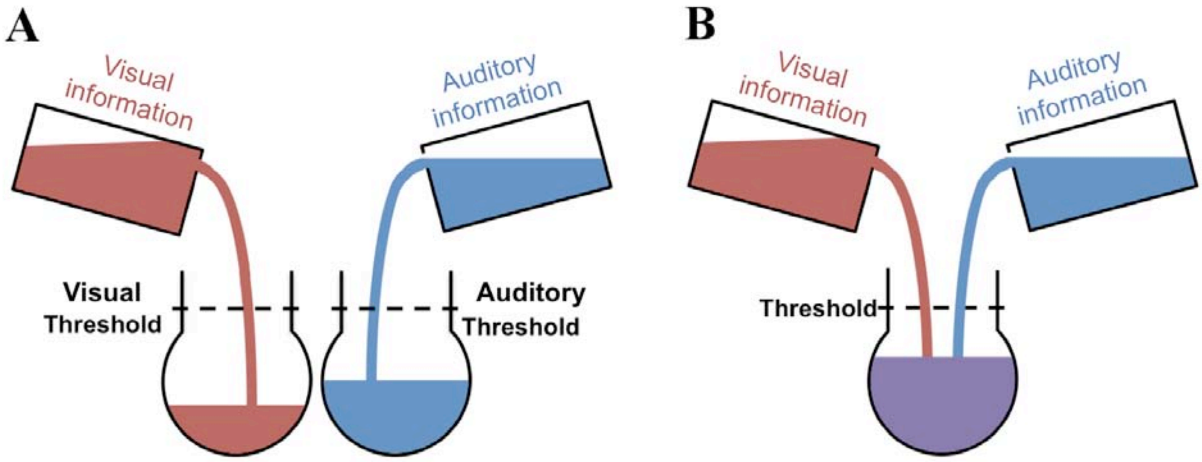

Figure 2.4. Illustration of the race and the co-activation models explanation of the faster behavioral response to redundant auditory-visual information.
Panel A. Illustration of the race model explanation. Evidences from the auditory and visual information processing are separately accumulated to meet the threshold for behavioral response. The bimodal response is as fast as the fastest unimodal response. The faster response with redundant auditory-visual information is due to statistical facilitation. Panel B. Illustration of the co-activation model explanation. Evidences from the auditory and visual information processing are combined to satisfy a single threshold for behavioral response. The bimodal response can be faster than the fastest response to unimodal information. The faster response with redundant auditory-visual information is due to multisensory integration.

Although the race model fully captured the RSE observed with redundant visual signals (Murray, Foxe, Higgins, Javitt, & Schroeder, 2001), it was often insufficient to account for the total RSE with auditory-visual redundant stimulation (e.g. Diederich & Colonius, 1987; Giray & Ulrich, 1993; Gondan, Lange, Rösler, & Röder, 2004; Miller, 1982). The race model was violated because the fastest reaction times in response to the multisensory redundant stimuli were statistically shorter than the fastest reaction times in response to a separate unimodal signal. This observation can be explained by the **co-activations model** (Miller, 1982). In this model, the brain combines the evidence from the processing of the different sensory signals in order to satisfy a single threshold and consequently lead to faster response with multisensory stimuli (see Figure 2.4B). This model captures the findings that multisensory stimuli can initiate a response at a time at which unimodal signals cannot induce a response by themselves. A RSE, which violates the Race model, suggest that the behavioral speed gain observed is linked to the integration of multisensory information.

RSE has been observed in different studies using detection tasks. For example, Miller (1982) asked their participants to detect an auditory bell stimulus and a visual asterisk stimulus when presented separately or synchronously. They found that detection of the auditory-visual stimulus is faster than the detection of each sensory signal alone. Furthermore, the responses were even too fast to be induced by a statistical facilitation suggesting that the RSE was, at least in part, linked to multisensory integration (Miller, 1982). Several studies found similar results with simple auditory and visual stimuli such as sound bursts, light flashes and geometric figures (Diederich & Colonius, 1987; Giray & Ulrich, 1993; Gondan et al., 2004). The RSE violating the Race model in detection tasks is not limited to auditory-visual redundant signals, it has also been demonstrated for visuo-tactile (Gondan et al., 2004), and auditory-tactile redundant signals (Gondan et al., 2004; Zampini, Torresan, Spence, & Murray, 2007). Further, Diederich and Colonius brought evidence for a facilitation of detection, higher than the Race model would predict, in response to trimodal events composed of flashes of light, simple tones and vibration on the toes (Diederich & Colonius, 2004). Their study also demonstrated an inverse effectiveness effect on reaction times. The RSE size increased as the intensity of the unimodal sensory stimuli decreased.

Multisensory information also facilitates object recognition. This has been demonstrated with experimental paradigms of the type Go-No Go, where participants are required to respond to sensory signals if, and only if, they come from a target object. Molholm and colleagues found evidence of a faster recognition of animals with auditory-visual information. They observed a RSE with a violation of the Race model when comparing reaction times to visual pictures of different animals (e.g. a cow) and to corresponding auditory vocalizations (e.g. cow lowing sound) presented individually to the synchronous presentation of both visual and auditory information (Molholm, Ritter, Javitt, & Foxe, 2004). The recognition of animals was facilitated by auditory-visual information in a way suggesting that multisensory integration is involved. Similar conclusions have been drawn for the recognition of simple auditory-visual events, consisting of noise burst and light flashes (Gondan, Niederhaus, Rösler, & Röder, 2005) or of a letter coupled to a pitch tone (Miller, 1991), and for the recognition of more complex events such as the coupling of a visual color presentation to the vocalization of the color designation (Laurienti et al., 2004). Moreover, Suied and colleagues (2009) supported the faster recognition of auditory-visual realistic objects with their study, in which they presented a frog and a phone in 3D using virtual reality techniques. They found a large RSE, coherent with an involvement of multisensory integration processes in the facilitation of recognition (Suied et al., 2009).

Beyond the facilitation of detection and recognition, multisensory information has been shown to facilitate visual searching. When searching for a cellular phone, one strategy that is often used is to call it in order to make it ring and/or vibrate. Empirical data support the fact that this is indeed an efficient strategy. Ngo and Spence (2010) delivered single tones stimuli and vibro-tactile stimuli to their participants while they were completing a visual search task. When searching for a visual target among distractors, a spatially coherent auditory or tactile signal substantially enhanced the performance of participants in terms of reaction time and accuracy (Ngo & Spence, 2010).

More accurate behavioral responses

Multisensory information also seems to serve a more accurate comprehension of the external events. The different senses have access to different information about the external environment. The different cues they capture are often complementary. For example, touch can provide information about the back of an object, which cannot be captured by vision (Newell, Ernst, Tjan, & Bülthoff, 2001). This complementarity between the different sensory cues can help disambiguate some situations as illustrated by the stream/bounce effect (Sekuler, Sekuler, & Lau, 1997). The stream/bounce video display represents two identical visual disks moving toward one other. At some point, the disks overlap and then pass each other. The situation is ambiguous: the disks could have either been streaming past each other or colliding and bouncing apart. A sound of collision at the time of disk overlap disambiguates the situation and leads to the perception of a bouncing movement.

Multisensory information has also been shown to improve speech comprehension. For instance, following and participating to a discussion during a social event like a cocktail party requires comprehending what your interlocutor is saying despite the ambient noise. Looking at the interlocutor facial and lips movements is one way to improve speech comprehension. A study examining the contribution of visual cues to speech perception in noisy situations empirically demonstrated that visual observation of the speaker increases speech intelligibility (Sumby & Pollack, 1954). Moreover, if listening to unfamiliar sentences, as can be the case in international events where the spoken language might not be your native language, the combination of visual and auditory signals of speech can radically improve comprehension. Using low-filtered sentences as unfamiliar sentences, Risberg and Lubker indeed found that the accuracy in word perception when both visual and auditory cues are provided is substantially higher than the accuracy when only individual sensory cue is provided. The gain is even superadditive compared to the accuracy measured in response to the unimodal cues (Risberg & Lubker, 1978).

How the combination of multisensory information leads to a more accurate comprehension of external events? It seems that the different sensory information do not have a similar influence on the production of an integrated, unified percept and one sensory cue often dominates the other. The ventriloquism effect is often cited as an example of the dominance of vision over audition during auditory-visual integration because the multisensory estimate of the speaker location tends rather toward the location of the visual cue than toward the location of the auditory cue. However, this dominance of vision over audition is not a general rule for multisensory integration. For example, the double flash illusion reveals a dominance of audition over vision: when a single flash of light is accompanied by multiple beep sounds, the observer perceives multiple flashes (Shams, Kamitani, & Shimojo, 2002).

The **Maximum Likelihood Estimation (MLE) model** is one of the models that have been put forward to account for how sensory cues are integrated. This model derives from Bayesian probability theory and proposes that the sensory cues are integrated in a statistically optimal manner meaning that the resulting multisensory estimated percept is most likely to be accurate. Each of the different sensory cues is processed by the nervous system and gives rise to an estimated percept, which is corrupted by noise. According to the MLE model, the multisensory estimate results from a linear combination of different sensory estimates weighted as a function of their reliability. The more reliable the sensory estimates, the higher the weight. The reliability is inversely related to the variance of the estimate, which is in turn linked to the noise corrupting the sensory estimate. The MLE model states that the multisensory estimate has the lowest variance possible, that is a variance lower than either variance of each sensory estimate alone, and is consequently the best estimate possible (see Figure 2.5 for an illustration of the model). In other words, when integrating an auditory and a visual spatial cue, for example, the reliability of the information we get from each cue would be taken into account to elaborate the multisensory percept more likely to be true.

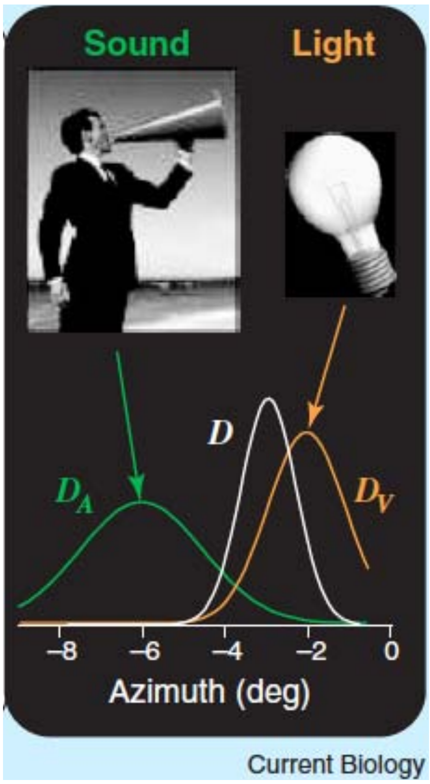

Figure 2.5. Maximum Likelihood Estimation of the spatial location of an auditory-visual event (adapted from Banks, 2004).
The green and yellow curves respectively represent the distribution of the auditory (D_A) and visual (D_V) estimated location of the event when the auditory cue is presented at -6 degrees and the visual cue is presented at -2 degrees. The variance of the D_A is higher than the one of D_V, indicating that the auditory estimate is less reliable than the visual estimate. The white curve represents the distribution of the auditory-visual (D) estimated location of the event if multisensory integration follows the Maximum Likelihood Estimation model. The pic of D is closer to the pic of D_V and the variance of D is smaller than both the variance of D_A and D_V, indicating that the auditory-visual estimated location is more reliable than the unisensory estimates.

The predictions of the MLE model have been confronted to empirical data and have been verified for many different situations (Ernst & Bülthoff, 2004), suggesting that sensory information is often integrated in a similar fashion to the one described by the MLE model. For example, Ernst and Banks (2002) investigated the integration of visual and haptic discordant cues about an object height by manipulating the degree of visual noise. Their study found that participants' height judgments were very similar to those predicted by the MLE model (Ernst & Banks, 2002), relying less on the visual cues if they are noisy.

The MLE currently is a very popular model because it can explain observed phenomena as well as more older ideas such as the "modality appropriateness hypothesis" (Welch & Warren, 1980). The modality appropriateness hypothesis relies on the fact that vision is more sensitive than audition for spatial judgment tasks whereas audition is more sensitive than vision for temporal judgment tasks and proposes that the integration of discrepant sensory cues leads to a multisensory percept biased toward the sensory modality that is more appropriate to the task at hand. Both models can explain the ventriloquism effect. However, unlike the modality appropriateness hypothesis, the MLE model also explains the phenomenon of "reverse ventriloquism", where audition dominates vision in a spatial judgment task if the visual cue estimate is very noisy (Alais & Burr, 2004).

Increased sensitivity for triggering behavioral responses

In addition to its effect on speed and accuracy of behavioral responses, multisensory information is also known as increasing detection sensitivity. Sub-threshold sensory signals, which are consequently not always detected when presented alone, can be detected more efficiently if combined with another sensory signal.

For instance, a bimodal event composed of a sub-threshold noise burst and a sub-threshold light flash is more easily detected when compared to the sub-threshold noise burst alone (Lovelace, Stein, & Wallace, 2003) or the sub-threshold light alone (Bolognini, Frassinetti, Serino, & Làdavas, 2005). Similar results have been obtained in a study examining a more realistic situation. In this study, the detection of being touched on the face was investigated. Participants received sub-threshold tactile stimulation on their face and could see a video of themselves being touched at the same time. Whereas being touched on the face was not always detected with the sub-threshold tactile cue, the presentation of both visual and tactile cues enhanced detection (Serino, Pizzoferrato, & Làdavas, 2008).

This gain in detection sensitivity for triggering behavioral responses with multisensory information resembles the response enhancement observed in the neurons of the cat superior colliculus. The multisensory response enhancement increases as unimodal stimulus effectiveness decreases so that stimuli, which are not capable of evoking any reliable response on their own, can induce a strong response when combined. Multisensory integration seems to aid the detection of an event when none of the individual sensory stimuli are effective enough to guarantee detection.

***

Put in a nutshell, multisensory information leads to behavioral gains. The combination of the sensory cues aids detection and allows faster and more accurate behavioral responses.

> Most events in the external world stimulate more than one of our senses. Research has shown that we effortlessly take advantage of this multiplicity of information to interpret and comprehend the environment in the most accurate and efficient manner. Multisensory information leads to perceptual and behavioral gains, which are, at least in part, linked to the combination and integration of sensory information at multiple stages of cerebral processing.

3. MULTISENSORY PROCESSING OF AFFECTIVE STIMULATION

3.1. Affect

Affective experiences are part of our daily life. They are involved in almost all of our interactions with our surrounding environment and we commonly use different terms to describe them: meeting with a friend was pleasant; a horror movie was frightening; a dish was really disgusting… These descriptions often come with details about our bodily states and behavioral reactions: "I was so happy that I could not stop smiling." or "I was really afraid when I saw this spider; I was startled and my heart started pounding in my chest.". The study of affect is critical to fully understand how we experience the world and interact with it. The research on affect has emerged for a while now and has met with some challenges. An empirical investigation necessitated a clear definition of the term "affect" as well as a definition of the different phenomena involved.

Definition of affect

Different theories have emerged to define affect but there is, at present, no generally accepted theoretical framework. In the current research work, I have adopted the definition of appraisalist theories of affect (Lazarus, 1991). The term affect is a global reference to all emotion-related processes in response to an emotional stimulus. This includes **emotional responses**, which embrace different phenomena:

- Autonomic responses such as heart rate, respiratory rate, blood pressure sweating (electrodermal activity) and pupil dilatation changes
- Neuroendocrine responses inducing changes in the concentration of different hormones in the blood
- Somatomotor responses including facial, gestural, vocal and behavioral changes

Affect also includes **conscious affective or emotional experiences**, also called feelings, which are the subjective experience of emotion-related changes in the central and peripheral nervous systems and **emotional regulation** processes, which allow producing emotional responses and experiences that are appropriated to the context. Moreover, in comparison to earlier theories of affect (Cannon, 1929; James, 1884), appraisalist theories emphasize the presence of a process, the **appraisal or identification of stimulus salience**, which may occur with or without awareness and precedes the other processes. These cognitive appraisal theories propose that affective processing consists of three stages (see Figure 3.1; Damasio, 1998; Phillips, Drevets, Rauch, & Lane, 2003):

(1) The appraisal and identification of the emotional significance of the stimulus

(2) The production of a specific affective state in response to the stimulus. This includes the elicitation of autonomic, neuroendocrine and somatomotor responses as well as conscious emotional experiences

(3) The regulation of the affective state and emotional behavior, which may involve a modulation of (1) and (2)

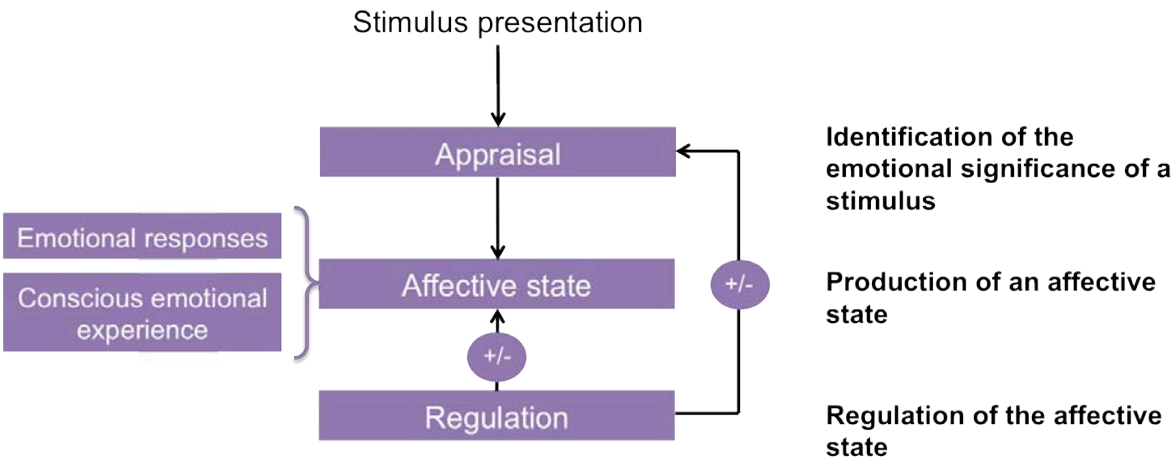

Figure 3.1. Affective processing stages according to the cognitive appraisal theories (adapted from Phillips et al., 2003).

The different components of affect can be experimentally addressed using one or a combination of different neurophysiological, physiological, behavioral and subjective reporting methods and techniques (Bradley & Lang, 2002). The process of appraisal and identification of the emotional significance of a stimulus can be referred to as **emotion perception** and can be investigated with behavioral and neurophysiological methods. The process of production of an affective state in response to a stimulus can be referred to as **emotion induction**. Emotional responses can be assessed with physiological and behavioral measures and the investigation of conscious emotional experience is measured with subjective reports.

Affect and survival

It is now well established that affect promotes survival. The ability to identify a potential threat or a potential good nourishment source and to rapidly implement appropriate behaviors is indeed advantageous in order to survive. Affect is generally considered as composed of different systems that have evolved to provide efficient solutions for the interaction with the environment (LeDoux, 1998).

One proposition, which I adopted in this thesis, assumes that affect is fundamentally organized around two motivational systems: a defensive system and an appetitive system (Bradley, Codispoti, Cuthbert, & Lang, 2001; Lang, Bradley, & Cuthbert, 1997). The defensive system principally is responsive to cues that threaten life and thus implements preservation behaviors such as withdrawal, escape and attack. On the other hand, the appetitive system is responsive to cues that promote life and produces a basic behavioral repertoire including approach, ingestion, copulation and care giving. According to this motivational approach, affect promotes survival by implementing contextually appropriate behaviors, with actions pushing us away from stimuli that could jeopardize our life and actions pulling us towards stimuli promoting life.

Characterization and classification of affective states

The different affective states that can be identified in daily life needed to be characterized and classified in order to investigate them in the laboratory. Two main types of classifications have been proposed: classifications into discrete categories and classifications along dimensional parameters (Lewis, Haviland-Jones, & Barrett, 2008).

The discrete classifications rely on theories called discrete or basic emotions theories. These theories propose that humans have evolved to have a limited set of basic affective states. Each of these affective states is unique and universal in its adaptive significance and expression. In other words, each affective state serves a specific evolutionary adaptation and conveys a specific set of emotional responses (endocrine responses, autonomic responses, facial expression, behavioral changes…), which do not vary from one individual to another. One famous attempt to characterize basic affective states is the work of Ekman and Friesen, who studied the universality of facial expressions (Ekman & Friesen, 1971). They travelled around the world to observe the facial expressions of individuals from different cultures, when they were happy, angry, sad, frightened, disgusted or surprised. They discovered that the facial expressions conveyed by these affective states were pretty much the same for all cultures. From this work, it was suggested that there were six basic affective states: happiness, anger, sadness, fear, disgust and surprise. However, to date there is no consensus on the number of basic affective states or on whether a list of categories is adequate to capture our full range of affective experiences.

Another approach is the dimensional characterization of affective experiences. The most commonly used is the characterization along two orthogonal dimensions: valence and arousal (Russell, 1980). The valence is a continuum specifying how positive or negative the affective experience is, whereas the arousal refers to the intensity of the experience. The affective experiences can then be placed within a two-dimensional space, where the abscissa axis represents the scale of valence and the ordinate axis represents the scale of arousal. As Figure 3.2 shows, the scale of valence range from negative or unpleasant to positive or pleasant and the scale of arousal range from low arousal or low intensity to high arousal/intensity. This classification is not optimal either to capture the totality of affective experiences.

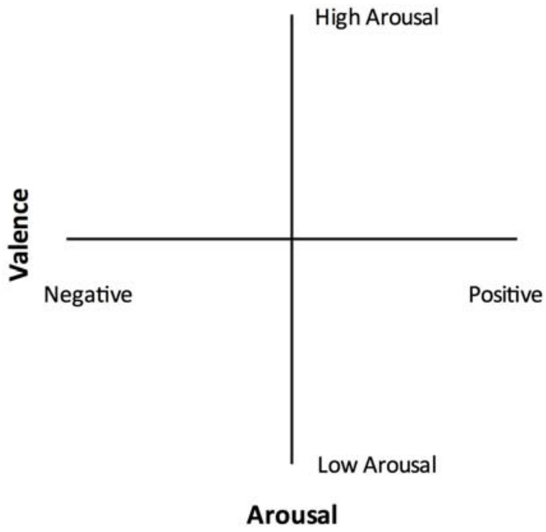

Figure 3.2. Affective two-dimensional space defined by valence and arousal scales. Different affective states can be placed within this space with highly negative or unpleasant states located in the upper left quadrant, low negative or unpleasant states in the lower left quadrant, highly positive or pleasant states in the upper right quadrant and low positive or pleasant affective states in the lower right quadrant.

Although both the discrete and dimensional approaches are not completely adequate to define and describe affect, they provide a good framework for research (Barrett, 1998). Moreover, the dimensional approach is not incompatible with the basic emotions approach given that each of the basic affective states can be placed in the affective two-dimensional space. In my work, I adopted the dimensional approach with the valence reflecting which motivational system is activated (defensive or appetitive) and the arousal indicating the intensity of motivational activation.

Affect and sensory information

Affective processing has primarily been investigated on single sensory modalities, mostly vision. Research on unimodal affect has brought evidence for a prioritization of perception of emotional stimuli (Brosch, Pourtois, & Sander, 2010), which is coherent with the relevance of affective cues for survival. Moreover, the investigation of affective processing has shown that the affective value of information is encoded from very early stages of cerebral processing (as early as 100 ms post stimulus; Olofsson, Nordin, Sequeira, & Polich, 2008). Affective processing enhances cerebral activity in many brain areas, including structures of the core affective neural network such as the amygdala as well as structures involved in sensory, attentional, mnesic or decision processes (Kober et al., 2008). While a unimodal focus allows for the establishment of fundamental knowledge on affective processing, the fact that affective information is perceived *via* multiple senses in a natural environment implies that a multisensory approach of affective processing is crucial to further understand human affect. Research on the processing of multisensory affective cues has emerged over the last 20 years. The following section provides an overview of the main findings.

3.2. Multisensory processing of affective stimuli

The number of studies investigating multisensory affective processing has increased in the past decades. This increase was spurred on by the explosion of research activity in multisensory processing. 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. Correctly identifying and reacting to affective events should enhance the chances of survival. The investigation of how emotional signals coming from different sensory modalities are combined is only beginning and started by examining emotional perception, i.e. the processes of identification and recognition of the affective states expressed by emotional stimuli. Studies have mostly focused on auditory-visual affective processing and have mainly used natural pairs of faces and voices conveying affective information (Klasen, Chen, & Mathiak, 2012). The following sections provide an overview of the current understanding of auditory-visual affective processing.

3.2.1. Auditory-visual integration of affective cues

In the same manner as spatial cues coming from different sensory modalities can be integrated to produce a unified percept of the spatial location of an event, it has been demonstrated that different sensory emotional cues can be combined to identify the affective state expressed by a stimulus.

In a study from 2000, De Gelder and Vroomen presented to their participants pairs of faces and voices with varying degrees of discordance between the affective state expressed by the face and by the tone of voice. They created different visual stimuli, which were pictures of faces from a morphed continuum between extreme happiness and extreme sadness (see Figure 3.3 for an example of a morphed continuum of facial expressions between happiness and sadness) and used sentences pronounced with sad or happy tones as auditory stimuli. For each couple of visual and auditory stimuli that were presented synchronously, participants had to pay attention only to the facial stimuli and indicate as quickly as possible whether they perceived that the person was feeling happy or sad.

Figure 3.3. Example of a morphed continuum of facial expressions from happiness to sadness (from Teunisse & De Gelder, 2001).

De Gelder and Vroomen collected the participants' responses for each facial expression ranging from extremely happy to extremely sad when presented with tone voices and compared them to participants' response when they were presented with the facial expression only. They observed an effect similar to the ventriloquism effect, "an emotional ventriloquism effect", when the facial expressions were ambiguous. Participants' judgment of the person's affective state shifted towards the affective cue conveyed by the unambiguous voice tone. When identifying the affective state of the person, an ambiguous facial expression presented with a sad tone voice lead more often to a sad judgment than when the face was presented alone. Conversely, the affective state was more often judged as happy with the happy voice tone. They also observed this "emotional ventriloquism effect" if participants had to pay attention to tone voices varying along a continuum from fearful to happy that were presented synchronously with fearful or happy faces. Whereas the "emotional ventriloquism effect" appeared with little discordance between sensory affective cues, the combination of the auditory and visual cues did not occur when the facial and vocal expression were clearly incongruent. These findings support the fact that affective cues coming from auditory and visual sensory modalities can be combined – providing that they are not incongruent— to produce a singular percept of the affective state of a stimulus (De Gelder & Vroomen, 2000). The same conclusion has been drawn from a study using computer-generated animated facial expressions synchronized with affective auditory speech (Massaro & Egan, 1996), from a study using whole body expression coupled with affective tone voice (Van den Stock, Righart, & De Gelder, 2007) and even from a study using non-natural pairs of whole body expression coupled with emotional music excerpts (Van den Stock, Peretz, Grèzes, & De Gelder, 2009).

The multisensory integration of affective cues has been assumed to be a mandatory and automatic process. This hypothesis is supported by different findings. First, even when instructed to only pay attention to one sensory modality when judging the affective state of a multisensory stimulus, participants' answers are influenced by affective cues delivered by the other sensory modality that they are instructed to ignore (De Gelder & Vroomen, 2000; Van den Stock et al., 2009, 2007). Secondly, the cross-modal influence is not constrained by attentional resources given that it occurs independently of the mental workload in dual-task situations (Vroomen, Driver, & de Gelder, 2001). Furthermore, cross-modal influences can also be observed in blind-sighted patients who are not aware of the visual affective cues of auditory-visual events (De Gelder, Morris, & Dolan, 2005). Moreover, a study examining cross-modal influences in an emotional categorization task brought evidence suggesting that the integration process occurs prior to response selection (Föcker, Gondan, & Röder, 2011).

* * *

The aforementioned findings demonstrate that multisensory affective information can be integrated to create a single percept of an event's emotional significance. The multisensory integration of affective cues follows rules similar to those governing the multisensory integration of neutral cues: the affective information coming from the auditory and visual modalities are effortlessly combined into a unified percept. Moreover, the semantic affective congruence is taken into account to combine cues that are meaningfully related.

3.2.2. Neural consequences of auditory-visual affective stimulation

Studies have mainly used natural pairs of faces and voices conveying affective information to investigate auditory-visual processing of affective stimuli. With this kind of complex stimuli, semantic congruence is an important factor for the integration of two sensory signals (Laurienti et al., 2004). Thus, two kinds of approaches have been used to examine the neural consequences of auditory-visual combination of affective information: (1) contrasting the neural response to auditory-visual affective stimulation with the response to the corresponding unimodal components and (2) contrasting the neural response to auditory-visual affective stimulation with emotionally-incongruent bimodal stimulation (see Figure 3.4).

The time-course of auditory-visual integration of affective cues

Electrophysiological techniques such as magnetoencephalography (MEG) and electroencephalography (EEG) have high temporal resolutions and enable the monitoring of the time-course of multisensory affective processing in the brain. Electrophysiological studies with auditory-visual affective stimuli have provided strong evidence of an early integration of multisensory information in the processing of affect.

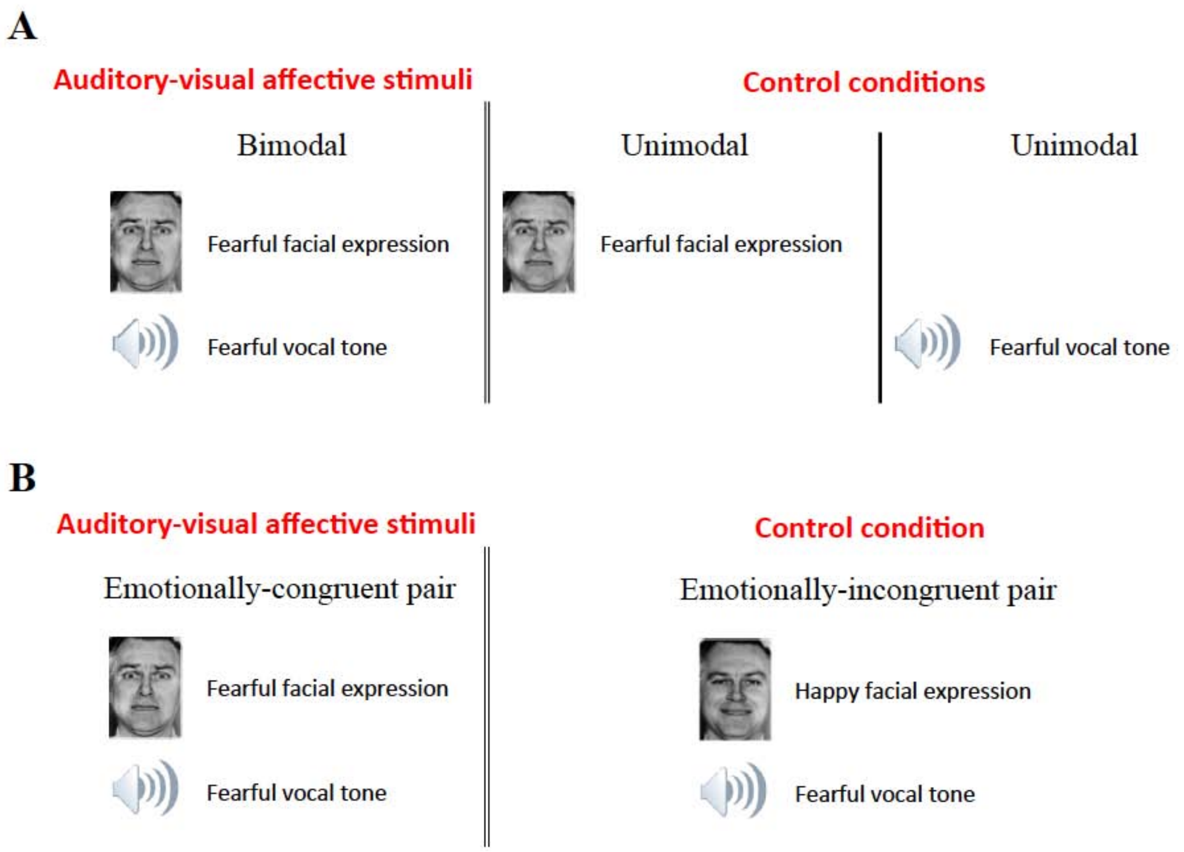

Figure 3.4. Main approaches used to investigate auditory-visual processing of affective stimuli (inspired by Pourtois et al., 2005 and Dolan et al., 2001).
Panel A. The neurophysiological responses to auditory-visual affective stimuli are contrasted with the responses to the corresponding unimodal affective stimuli. Panel B. The responses to auditory-visual affective stimuli are contrasted with the responses to incongruent auditory-visual affective stimuli.

Using the event related potential (ERP) technique, which is derived from EEG, De Gelder et al. showed that a facial expression combined with an emotionally incongruent affective voice evokes an early mismatch negativity response around 180ms post-stimulus (De Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999). This mismatch negativity is known as an indicator of a deviation between the perceived stimulus and the expectation (Garrido, Kilner, Stephan, & Friston, 2009). The fact that it is observed at 180ms after the presentation of emotionally incongruent auditory-visual face-voice pair strongly suggests that an interaction between the different sensory cues has already happened at this stage of processing. Moreover, in this study, participants were instructed to ignore the auditory stimulus. Thus, the mismatch negativity response supports a mandatory and automatic integration of the auditory and visual affective cues.

With the same ERPs technique, Pourtois et al. found evidence of an even earlier combination of auditory and visual affective cues. They presented their participants with faces and voices expressing anger or sadness coupled either into an emotionally congruent bimodal stimulation or coupled into an emotionally incongruent bimodal stimulation. They observed a modulation of the amplitude of the early sensory components N1 and P2 in response to emotionally congruent stimuli when compared to incongruent ones, suggesting a combination of the two affective signals as early as 110ms post-stimulation (Pourtois, De Gelder, Vroomen, Rossion, & Crommelinck, 2000). In a subsequent study, emotionally congruent pairs of faces and voices were also found to modulate the latency of the P2b component. This component, consisting of a positive wave occurring around 240ms post-stimulation and reflecting cerebral activity in posterior brain areas, was observed earlier in response to emotionally congruent compared to emotionally incongruent pairs of face and voice. This suggests that auditory-visual incongruent pairs, in term of emotional expression, delay the processing of information (Pourtois, Debatisse, Despland, & De Gelder, 2002).

Another study also used affective vocalizations but instead of coupling them with faces, they coupled them with dynamic whole-body affective expressions (Jessen & Kotz, 2011). They compared EEG signals in response to auditory-visual stimuli to the response with only unimodal auditory stimuli and found a modulation of the auditory N1 component amplitude followed by modulation of the P2 component amplitude, suggesting an early influence (around 100ms post-stimulus) of visual cues on auditory cues processing.

Whereas the previously mentioned studies employed natural pairs of auditory-visual affective stimuli, one study used “arbitrary” pairs. Unlike natural pairs, “arbitrary” pairs do not usually co-occur in natural environments (De Gelder & Bertelson, 2003). Spreckelmeyer et al. presented their participants with emotional scenes coupled with sung notes as bimodal stimuli. They showed that emotionally congruent pairing of happy pictures and happy sung notes led to a modulation of the P2 component compared to incongruent stimuli (Spreckelmeyer, Kutas, Urbach, Altenmüller, & Münte, 2006), demonstrating an early integration of auditory-visual affective cues even with non-natural pairs of auditory and visual affective stimuli.

Instead of ERPs, Hagan et al. used MEG and time-frequency analyses to examine the processing of auditory-visual affective stimuli. They compared the response to affective face pictures and voices expressing fear with the response to the individual unimodal stimuli. They observed a superadditive response to bimodal stimuli expressing fear in the broadband (3-80Hz) within the first 250ms post-stimulation (Hagan et al., 2009). The response to the auditory-visual stimuli was higher than the sum of the responses to the visual and auditory stimuli presented on their own. They located the source in the posterior superior temporal sulcus (pSTS), suggesting that the pSTS would have a role in the integration of affective cues. However, a recent study also found a superadditivity in the pSTS, although delayed, in response to incongruent affective auditory-visual cues, suggesting that pSTS would rather be involved in the resolution of auditory-visual affective cues than in their integration (Hagan, Woods, Johnson, Green, & Young, 2013).

In their MEG study, Chen and colleagues used dynamic faces and voices expressing angry or happy affective states and also found an early modulation of cerebral activity (from 100ms post-stimulus) in response to auditory-visual affective stimuli compared to the corresponding affective unimodal stimuli (Chen et al., 2010). They localized the sources in frontal areas and thalamus.

The cerebral sites of auditory-visual integration of affective cues

Neuroimaging techniques such as functional magnetic resonance imaging (fMRI) or positron emission tomography (PET) are complementary to electrophysiological techniques. Whereas the latter offer a high temporal resolution, neuroimaging techniques provide high spatial resolution allowing for the investigation of the cerebral areas involved in the integration of auditory-visual affective cues. Several cerebral sites have been identified as involved in the multisensory processing of affective stimuli. Particularly, compelling evidence of the involvement of superior and middle temporal cortical areas in the integration of affective auditory-visual cues has been brought to light.

When contrasting cerebral activity in response to bimodal affective stimuli with cerebral activity in response to the unimodal affective cues presented on their own, a series of fMRI studies have found a stronger activation in the superior temporal structures (around the pSTS) with pictures of facial expression coupled with affective voices (Park et al., 2010) as well as with video clip of dynamic facial expression coupled with affective voices (Kreifelts, Ethofer, Grodd, Erb, & Wildgruber, 2007; Li et al., 2013; Robins, Hunyadi, & Schultz, 2009). A role of the temporal lobe in the integration of auditory-visual affective cues was also demonstrated by a PET study (Pourtois, De Gelder, Bol, & Crommelinck, 2005). Affective still faces and voices evoked an enhanced response in the middle temporal gyrus (MTG) compared to either of the sensory affective cues in isolation. These findings, suggesting the involvement of superior and middle temporal structures in multisensory affective processing, are coherent with the fact that the STS is known as supporting multisensory integration of auditory and visual cues (see chapter 1).

Other brain structures have also been associated with the integration process of affective information. These structures include the thalamus (Ethofer, Pourtois, & Wildgruber, 2006; Kreifelts et al., 2007; Park et al., 2010), the amygdala (Park et al., 2010), the ventral posterior cingulate cortex (vPCC; Klasen, Kenworthy, Mathiak, Kircher, & Mathiak, 2011) and the insular cortex (Ethofer, Pourtois, et al., 2006), which are known to be involved in affective processing (Kober et al., 2008).

For now, however, only the superior temporal region, the amygdala and the vPCC seem to be involved in the specific integration of affective semantic information. So far, these are the only structures, which show a modulation of their activity as a function of the emotional congruency between the auditory and visual components of affective bimodal stimuli.

The modulation of activity in the superior temporal brain area was demonstrated in a study where the bimodal stimuli were non-natural pairings of faces and instrumental music pieces (Jeong et al., 2011). The faces and musical excerpts could both express happy or sad affective states or be coupled in an incongruent manner. A greater activation within the superior temporal gyrus (STG) for emotionally congruent pairs was observed as compared to incongruent ones, suggesting that superior temporal brain areas are involved in the integration of the affective semantic information of auditory-visual stimuli.

The modulation of the amygdala activity was demonstrated in a study where the bimodal stimuli were natural pairs of voices and still faces (Dolan, Morris, & De Gelder, 2001). The voices and faces expressed fear or happy states and were coupled in either an emotionally congruent or incongruent manner. Participants' cerebral activity in response to the bimodal stimuli was measured with fMRI while they were asked to judge the affective state conveyed by the facial expression and to ignore the voice. In comparison to incongruent bimodal stimuli, emotionally congruent pairs of face and voice lead to an increase of cerebral response in the left amygdala. This effect was observed even though participants were instructed to ignore the auditory affective cues, suggesting that an integration of affective semantic cues occurs in the amygdala regardless of the attentional focus. The implication of the amygdala in multisensory affective processing is further supported by the fact that activity decreases if an affective facial expression is coupled with a neutral voice or if an neutral face is coupled with a voice expressing an affective state (Müller et al., 2011).

Another study reported a modulation of amygdala activity as a function of the emotional congruency of bimodal stimuli and demonstrated that the vPPC is also involved in the integration of affective semantic cues (Klasen et al., 2011). They used video clips of dynamic faces coupled with voices both expressing angry or happy affective states in a congruent or incongruent manner. The emotionally congruent bimodal stimuli evoked a higher activity in the vPPC when compared to the incongruent bimodal stimuli.

* * *

Altogether, it seems that the different sensory cues of auditory-visual affective stimuli interact from very early stages of cerebral processing and in different brain structures (see Figure 3.5). As proposed by a recent model for multisensory affective processing (Klasen et al., 2012), two integration processes may occur: an integration of the physical features of the sensory cues in the thalamus, primary sensory cortices and superior and middle temporal brain areas and an integration of the affective semantic information in higher association cortices such as the vPPC and the STS/STG and mediated by the amygdala.

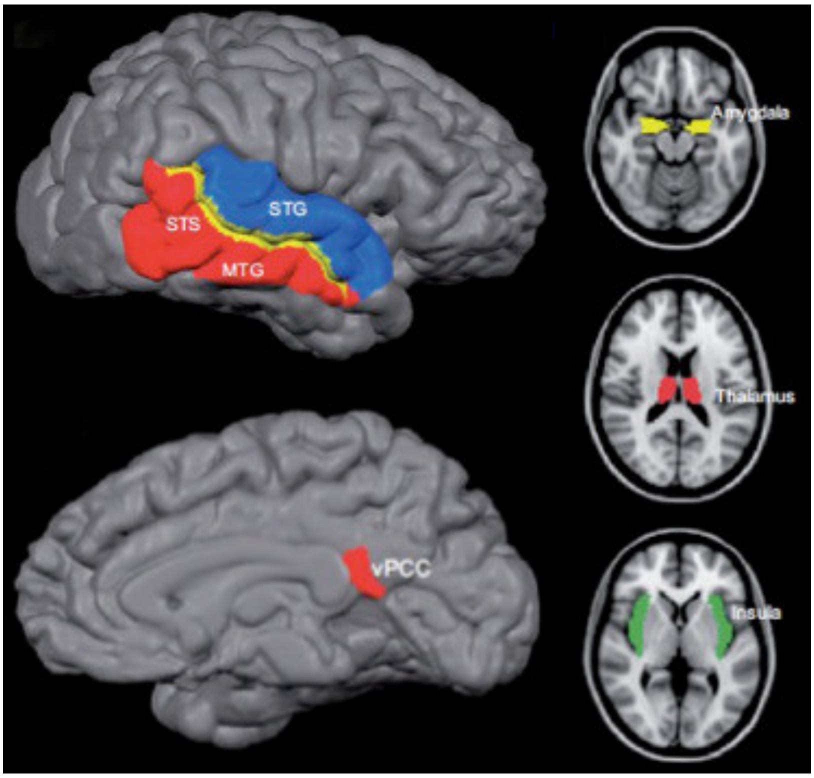

Figure 3.5. Brain structures involved in the integration of multisensory affective information (adapted from Klasen et al., 2012).
Temporal regions (top left) including the superior temporal gyrus (STG), the superior temporal sulcus (STS), the middle temporal gyrus (MTG) as well as the amygdala (top right) and the ventral region of the posterior cingulate cortex (vPCC; bottom left) emerged from neuroimaging studies as integration sites for semantic affective cues. The thalamus (middle right) and the insula (bottom right) seems to be involved in the combination of multisensory affective information but are not sensitive to affective semantic congruence.

3.2.3. Consequences of auditory-visual affective stimulation on emotion identification

It is now well established that redundant affective cues coming from the auditory and visual modalities facilitates the identification of the affective state conveyed by bimodal events. This effect of the combination of auditory and visual cues on the identification of events' emotional significance has been demonstrated with categorization tasks, where participants had to choose between two or several affective states which one corresponds to the affective state conveyed by a stimulus.

Faster identification of events' emotional significance

The categorization of an affective state expressed by a person is faster with redundant auditory-visual affective stimulation than with facial expression or vocal tone alone (Collignon et al., 2008; Li et al., 2013; Pourtois et al., 2005). However, a semantic emotional incongruence between the affective information respectively conveyed by face and voice slows the response (De Gelder et al., 1999; De Gelder & Vroomen, 2000; Dolan et al., 2001; Föcker et al., 2011). These results demonstrate that a redundant signal effect not only occurs with the detection and recognition of neutral objects, but also appears when identifying an emotional state. Moreover, the fact that emotionally incongruent information slows categorization demonstrates that the gain in speed is linked to the integration of affective semantic cues delivered by the visual and auditory modalities.

A study from Collignon and colleagues also supported the implication of multisensory integration in the faster behavioral responses observed with multisensory affective stimuli. Their participants had to categorize the affective state of a person based on either the bimodal presentation of voices and dynamic faces or based on unimodal stimulations. Participants' reaction times show a redundant signal effect, which violated the race model (Collignon et al., 2008). This means that the rapidity of their responses in the bimodal conditions can be explained by an integration of the auditory and visual cues. Furthermore, the increase of affective categorization speed found when both the facial expression and the vocal tone were presented was greater if the unimodal affective cues were noisy. This is coherent with the inverse effectiveness principle, which has been established as a key principle of multisensory integration (see chapitre 1).

More accurate identification of events' emotional significance

The categorization performance of events' emotional significance is also improved in terms of accuracy with redundant auditory-visual affective stimuli. The categorization of a person's affective state when both affective cues from the face and the voice are provided was found to be more accurate than when only the facial expression or the vocal tone is presented alone (Collignon et al., 2008; Kreifelts et al., 2007; Li et al., 2013). These results demonstrate that multisensory information not only improve the accuracy of the neutral objects recognition, but also lead to a higher accuracy when identifying an emotional state. Moreover, the accuracy in the recognition of the affective state conveyed by a facial or whole body expression is impaired if the vocal tone conveys an incongruent affective cue (Föcker et al., 2011; Tanaka et al., 2010; Van den Stock, Grèzes, & De Gelder, 2008) and the performance becomes even worse than with only visual cues (Van den Stock et al., 2008). This impairment suggests that the improved accuracy is related to the integration of semantic affective cues. This combination of the different sensory cues leads to a more robust and accurate emotional percept.

It seems that the different sensory affective cues are combined in a manner similar to sensory neutral cues. The "emotional ventriloquism effect" found by De Gelder and Vroomen grew with the ambiguity of the affective cue in the attended modality. The more the affective cue was unclear in the attended modality, be it auditory or visual, the more participants relied on the other cue to classify the affective state of the person presented to them. This is coherent with the maximum likelihood estimation model (see chapter 1), which assumes that the most reliable sensory cue is given the most weight during multisensory integration in order to produce the most reliable percept. This phenomenon was also observed when affective cues were ambiguous due to environmental noise. In a study using video clips of real dynamic faces and non-linguistic vocalization expressing fear or disgust, the reliability of the visual affective cues was modified with white noise. With this manipulation, the stimulation simulated a context where emotional perception takes place in a dark environment rather than a context where the affective expression is itself ambiguous. The results of this study showed that when judging a person's affective state, as the reliability of visual affective cues diminished, participants relied more on auditory affective cues (Collignon et al., 2008). This further supports the idea that multisensory affective information is integrated in an optimal manner to arrive to a singular, robust and accurate percept of events' emotional significance.

Beyond the facilitation of the identification and recognition of the valence of affective stimulus, multisensory affective information has been shown to increase the perceived intensity of an affective state. For example, we judge a person as being more afraid when we both see his/her face and hear his/her vocal tone expressing fear in comparison to when we are provided with only one of the sensory affective cues (Ethofer, Anders, et al., 2006; Müller et al., 2011). This effect was also found for happiness, sadness and anger states conveyed by facial expression and vocal tone (Föcker et al., 2011; Jeong et al., 2011). Furthermore, the affective state of bimodal stimuli with emotional incongruences between the facial and vocal cues is perceived as less intense than with emotionally congruent bimodal stimuli or with only unimodal stimuli (Föcker et al., 2011), suggesting a multisensory integration of affective semantic cues in the evaluation of the affective state intensity conveyed by a bimodal stimulus.

* * *

Put in a nutshell, the integration of auditory-visual affective information leads to a perceptual gain. The emotional significance of affective auditory-visual events are identified and evaluated more rapidly and more accurately. Moreover, the affective state expressed by auditory-visual events is also evaluated as more intense.

3.2.4. Consequences of auditory-visual affective stimulation on the induction of an affective state

As described in the previous sections, the investigation of multisensory affective processing has mainly focused on the perception of emotion. Currently, most of the studies have used affective categorization tasks to examine how multisensory stimulations influence the first stage of affective processing, i.e. the appraisal and identification of the emotional significance of stimuli. However, the impact of multisensory information on emotion induction, i.e. the production of an affective state in the perceiver, remains relatively undiscovered.

A couple of studies have investigated autonomic emotional responses induced by auditory-visual affective stimuli. Chapados & Levitin explored the electrodermal response induced by aesthetic music stimuli. They measured the electrodermal response induced by a musical performance when participants could both see and hear the performer and when they could only see or only hear him. They found that the participants' electrodermal responses for auditory-visual performances were higher than the sum of the electrodermal responses for each unimodal performances (Chapados & Levitin, 2008). This superaddivity is similar to the phenomenon observed during multisensory integration with neurophysiological measures and suggest that the integration of the affective visual and auditory cues of the performance is involved in the enhancement of the emotional response induced by the bimodal performance. Brouwer and colleagues found contrasting effects in their study. They measured heart rate and electrodermal response induced by affective auditory-visual stimuli or by the corresponding unimodal stimuli and did not find any effect of stimulus modality on autonomic emotional responses (Brouwer, Van Wouwe, Mühl, Van Erp, & Toet, 2013). However, they used non-natural pairs of affective pictures and sounds from the International Affective Pictures System

(IAPS) and the International Affective Digitized Sound (IADS) databases. Even though these bimodal stimuli were emotionally congruent, their semantic congruence was not really optimal and thus, the pictures and sounds could have been interpreted and processed as two different events. This could explain the absence of effect on physiological responses.

Some studies have explored the conscious emotional experience (feeling) induced by affective auditory-visual stimuli. Vines and colleagues examined the subjective aesthetic experience induced by musical performance. They asked their participants to report the intensity of their aesthetic experience when they could both see and hear the performer (see an example of performance in Figure 3.6) and in a unimodal condition when they could only hear the performer. They found that participants' aesthetic experience was more intense with bimodal performances (Vines, Krumhansl, Wanderley, Dalca, & Levitin, 2011; Vines, Krumhansl, Wanderley, & Levitin, 2006).

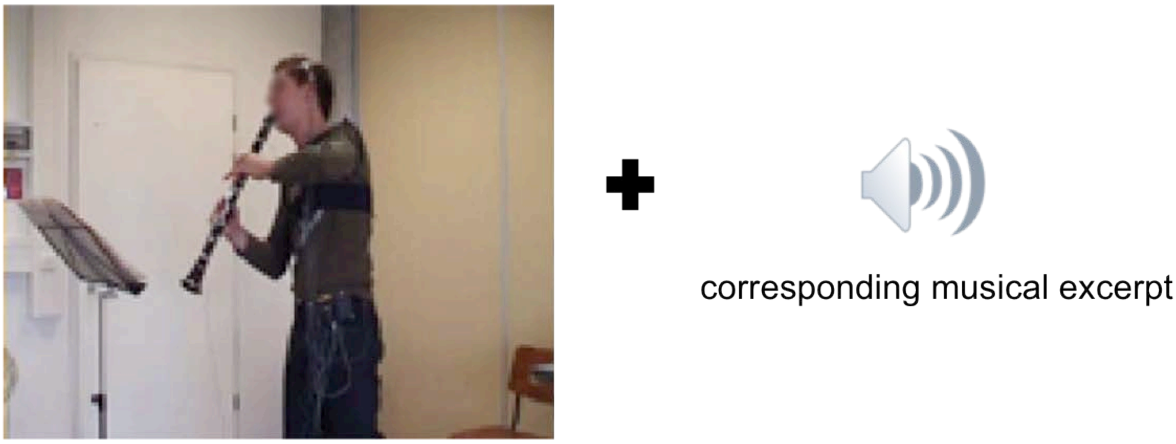

Figure 3.6. Example of the naturally multisensory stimuli used in the study of Vines et al., 2006 and 2011
They used video recordings of musical performances, which are natural auditory-visual stimulation. Participants could both see and hear the performance or only hear it.

In another study, Baumgartner and colleagues used non-natural pairing of affective pictures and affective music excerpts to induce feelings (see Figure 3.7 for an example of the type of stimuli they used). Their participants reported increased affective experience in the bimodal condition when compared to the condition wherein they only saw the affective pictures (Baumgartner, Lutz, Schmidt, & Jäncke, 2006). Even though the semantic congruency is not optimal with such non-natural auditory-visual pairs, the combined presentation of auditory and visual affective stimuli amplified subjective feelings.

Figure 3.7. Example of the type of non-natural pairs of visual and auditory stimuli used in the study of Baumgartner et al., 2006
They used non-natural pairs of pictures from the IAPS and musical excerpt. The pictures were fear-, sad- and happy- inducing pictures and they all contain humans or human faces. The affective pictures were presented alone or coupled with classical orchestral pieces evoking fear, sadness or happiness.

* * *

Studies investigating the consequences of multisensory affective stimulation on the affective state elicited in the perceiver remains sparse for the moment. The few studies, which have addressed this issue, mainly focused on the affective state induced by aesthetic musical stimulation or used non-natural pairs of affective stimuli. The findings revealed contrasting result on emotional responses and suggested that multisensory affective stimuli may amplify conscious emotional experiences.

Affective events often deliver cues across multiple sensory modalities. The investigation of multisensory processing of affective stimuli suggests that this availability of redundant affective cues leads to a gain in emotion perception, i.e. in the identification of stimuli's emotional significance. This gain seems to be, at least partly, due to the integration of the sensory affective cues at multiple stages of cerebral processing and in different brain structures (mainly the amygdala, the ventral posterior cingulate cortex and the superior temporal areas).

However, the influence of multisensory affective information on emotional induction, i.e. on the production of an affective state in response to an affective stimulus, remains relatively undiscovered. Very few studies have examined the affective state induced by multisensory affective stimuli. It is possible that the multiplicity of sensory affective cues amplify the emotional responses and the conscious emotional experiences.

4. SPACE AND AFFECT

We evolve in a three dimensional world. When perceiving objects and events occurring in our surrounding environment, one important feature that we extract and interpret is their spatial location. The spatial location can be represented within either an allocentric or egocentric referential frame (Klatzky, 1998). The allocentric representation refers to points in space external to the perceiver whereas the egocentric representation uses the perceiver as the referencing point. For example, when perceiving a dog, we can represent the dog as located at 30cm from the couch (allocentric representation) and as located at 2m from us (egocentric representation). The egocentric representation of the location of external objects and events is particularly relevant for affective events, which can promote or threaten survival. For example, in order to preserve our life in the presence of an aggressive dog, information about the location of the dog in relation to our body is critical in order to implement pertinent behaviors. This dog represents a potentially higher threat to our life if it is located at 50cm from our body than if it is located at 10m from us. The location of stimuli in the space around the body seems, thus, of substantial importance for affective processes. Additionally, research suggests that the egocentric representation of space is not a unitary construct. Events located at close distances to the body may be differently represented from the ones located far from the body (Rizzolatti, Fadiga, Fogassi, & Gallese, 1997). The next section describes the research from which this division of space around the body has emerged.

4.1. The space around us

An important milestone in research concerning the representation of space was Hall's proxemic framework (Hall, 1963, 1966). On the basis of the observation of humans use of space and of human social interactions, he proposed a four-tier organization of the space around the body: intimate space, personal space, social space and public space (see Figure 4.1). These different zones reflect the different distances used during social interaction according to the relationship between the interacting individuals. These distances decrease as the interacting individuals have closer relationships; conjointly, the amount of sensory information exchanged increases. Individuals who have a really close relationship such as romantic partners or members of a same family use an intimate distance. This is the distance used during comforting, wrestling or love-making behaviour. Insofar as the personal distance is concerned, there is no physical contact. This is a distance commonly accepted by two individuals who already know each other and interact regularly. At social distances, no contact would be possible without a modification of the individuals' locations. Public distance is used in public occasions such as a public performances or presentations.

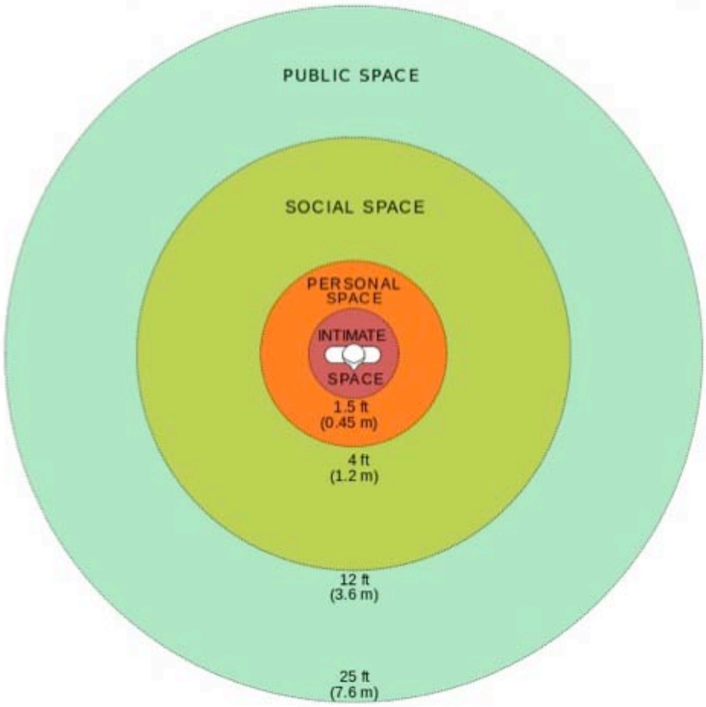

Figure 4.1. Illustration of Hall's proxemics framework.
This framework is based on the observation of the distance that American individuals adopt during social interactions in function of their relationship to the other person. The space around the body is organized in four different spatial spaces: intimate space, personal space, social space and public space.

Dichotomous representation of external space and affect

Subsequent work in the field of social psychology has mainly focused on personal space. The concept of personal distance comes from ethological studies, characterized as the distance that animals of non-contact species naturally keep between them (Hediger, 1955 as quoted by Hall, 1966). Hediger explained that these animals naturally behave as if an invisible bubble surrounded them. An intrusion into this bubble triggers a flight behavior in the animal and, if flight is not possible, the animal will attack the intruder.

Along with Hall, Sommer adopted this idea of a protective bubble surrounding the body to describe the personal space in humans. He defined personal space as “an area with invisible boundaries surrounding a person’s body into which intruders may not come” (Sommer, 1959). Although the bubble shape of personal space has been challenged (Hayduk, 1981a), research in social psychology has generally supported Hall and Sommer’s idea. This field of research focused on human use of space during social interactions and the term *interpersonal distance* was often employed to refer to personal space.

Differences in the adopted interpersonal distance were observed as a function of age, gender, culture and the relationship between the interacting individuals (see Aiello, 1987 or Hayduk, 1983 for a review). An inappropriate distance generally results in negative affective states for the individuals who feel that their personal space has been intruded upon (Aiello, 1987; Hayduk, 1978). Spatial intrusions are typically associated with the emergence of feelings of discomfort as well as with an increase of individuals’ autonomic activity; they additionally lead to compensatory behaviors such as withdrawal, flight or attempts to decrease the amount of sensory information provided by the intruder (by reducing eye contact for example). Consequently, the definition of personal space for humans has evolved to include an affective dimension. Hayduk proposed the following definition: “personal space is the area individual humans actively maintain around themselves into which others cannot intrude without arousing discomfort” (Hayduk, 1978). Moreover, the different theories that have been put forward to explain how the appropriate interpersonal distance is implemented, all agree on a crucial involvement of motivational appetitive and defensive systems (Aiello, 1987; Hayduk, 1983).

Recent findings have further supported the importance of affect in the implementation of personal space. Kennedy and colleagues demonstrated that an unaffected amygdala is crucial and even indispensable for the implementation of appropriate interpersonal distances during social interaction (Kennedy, Gläscher, Tyszka, & Adolphs, 2009).

The field has commonly assumed that affective processes drive the implementation of the personal space, at least partially. Therefore, the measure of personal space boundaries has often implicated affective measures such as physiological measure of autonomic arousal or subjective reports of affective experience. For example, research on personal space has often used stop-distance paradigms. In stop-distance paradigms, participants approach or are approached by another person and are instructed to stop the approach as soon as they start to experience discomfort. The distance at which the approach is stopped is considered as the personal distance. This empirical method has been used to assess the size and permeability of personal space (Hayduk, 1981b), the latter of which can be described by the relation between the extent of personal space intrusion and the level of experienced discomfort.

The idea that the representation of the location of events occurring in the surrounding environment is different according to the events' distance with the perceiver has emerged from this body of research on human spatial behaviors during social interactions. The marked discontinuity of affective responses brings evidence for a dichotomous representation of the location of external events at far or close personal distances. The internal egocentric representation of space seems to involve a dichotomous representation for space within or outside the emotionally-implemented personal distance.

Dichotomous representation of external space and sensory processing

More recently, additional evidence for a dichotomous representation of space has emerged from the field of cognitive neuroscience. In this field, the space immediately surrounding the body has been called *peri-personal space* (Rizzolatti et al., 1997) whereas the space far from the body has been called *extra-personal space*. Studies conducted with human patients have revealed that brain damage can specifically lead to impaired spatial awareness (neglect) in one of these spaces without neglect in the other one. Using a line bisection task, Halligan and Marshall observed that their patient showed a neglect in peri-personal space but not in extra-personal space (Halligan & Marshall, 1991). Conversely, Cowey and colleagues reported the cases of brain-damaged patients who showed a severe neglect in extra-personal space compared to peri-personal space (Cowey, Small, & Ellis, 1994). These findings support the idea that peri- and extra-personal spaces are distinctly represented in the brain. Further studies on both humans and non-human primates have suggested that the space near the body is represented differently in the brain from the space far from the body.

Findings from neurophysiological studies in monkeys have supported the idea of a neural circuit dedicated to the representation of peri-personal space. Using single-unit recordings, these studies have revealed neurons specialized in the coding of visual events occurring in peri-personal space. These neurons were identified in different areas of the monkey brain including the area F4 of the ventral premotor cortex (Gentilucci et al., 1988), the ventral intraparietal areas (Colby, Duhamel, & Goldberg, 1993), the parietal areas 7b (Graziano & Gross, 1995) and the putamen (Graziano & Gross, 1993). These neurons are bimodal neurons, responding to both tactile and visual events. They have tactile receptive fields on the surface of the body and visual receptive fields close to the tactile receptive fields (see Figure 4.2). A visual event located near the surface of the body triggers high responses (high firing rate) in these neurons, whereas the same visual event triggers weak responses when located far from the skin (Brozzoli, Makin, Cardinali, Holmes, & Farnè, 2012).

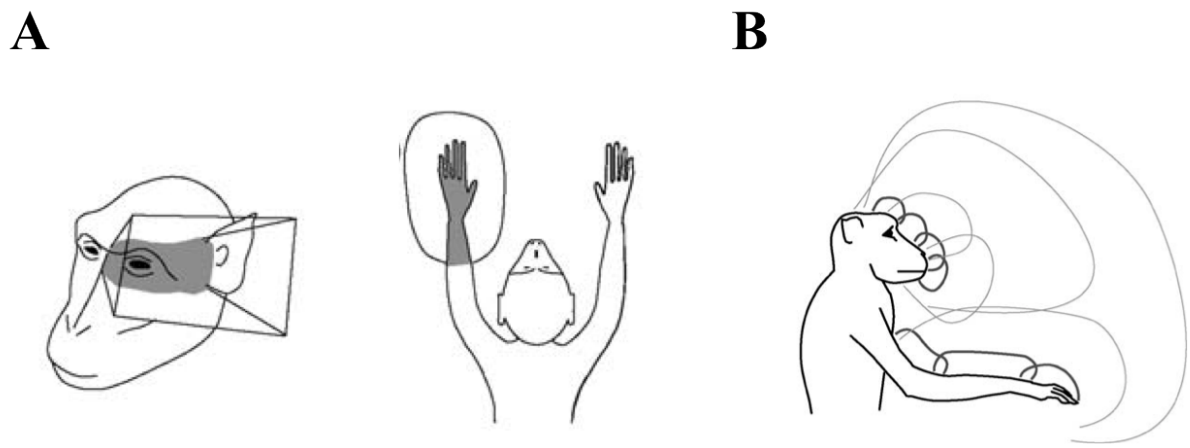

Figure 4.2. Peri-personal space representation in the monkey (from Graziano et al., 2006).
Panel A. Examples of tactile (shaded) and visual (boxed) receptive fields of two neurons of the monkey brain, involved in peri-personal space representation. Panel B. The different receptive fields of the multiple neurons allow the representation of peri-personal space.

The existence of a similar multisensory representation of the space surrounding the body in the human brain has been supported by several studies in healthy individuals and brain-damaged patients (Farnè & Làdavas, 2002; Làdavas & Farnè, 2004). In healthy individuals, evidence of a multisensory representation of peri-personal space has been brought to light with the crossmodal congruency task. During this task, participants are required to discern the location of a tactile stimulus on their hand between two possibilities, while trying to ignore a visual stimulus delivered at a congruent or incongruent location. The incongruent conditions lead to impaired accuracy and slower reaction times. This effect of incongruence was found to be especially large when the visual stimulus was presented near the tactile stimulus compared to when it was presented far from the tactile stimulus (Spence, Pavani, & Driver, 2004; Spence, Pavani, Maravita, & Holmes, 2004). Incongruent visual events located near the stimulated hand was an information more conflicting with the tactile stimuli than incongruent visual events far from the stimulated hand, suggesting the presence of an integrative system that monitors both the visual and tactile events near the body, i.e. within the peri-personal space. A similar conclusion was drawn from a study using auditory events instead of visual events with neurological patients (Farnè & Làdavas, 2002).

Research has sought to identify the cerebral structures underlying the representation of peri-personal space in the human brain; neuroimaging studies have revealed brain areas in the ventral premotor cortex (vPMC) and in the posterior parietal cortex (PPC) as serious candidates. An increased activity in these brain structures was found in response to both unimodal visual, auditory and tactile stimuli presented on and near the head (Bremmer et al., 2001). Tactile and visual stimuli on and near the face evoke responses in aligned maps in the ventral part of the intraparietal sulcus in the PPC (Sereno & Huang, 2006). Moreover, a modulation of the cerebral activity within vPMC and PPC was observed in response to a visual stimulus near the hand (Makin, Holmes, & Zohary, 2007). Furthermore, Serino and colleagues demonstrated that virtual lesions in the vPMC and PPC induced by repetitive transcranial magnetic stimulation (rTMS) cause a disruption of auditory-tactile interactions around the hand, suggesting that these brain areas are necessary for the multisensory representation of peri-personal space (Serino, Canzoneri, & Avenanti, 2011). These brain areas nicely correspond to areas identified in the monkey brain. In the human brain, a fronto-parietal network seems to specifically represent the space near our body.

* * *

Together, research in the fields of social psychology and cognitive neuroscience has brought evidence of a dichotomous representation of the external space. The clear discontinuities of affective responses and sensory processes seem to reflect a same division of external space representation in two spaces. The space near the body, i.e. personal or peri-personal space, is represented differently from the space far from the body, i.e. the extra-personal space. For the rest of this manuscript, I will use the term peri-personal space (PPS) to refer to the space near the body. The specific representation of PPS is coherent with the fact that events occurring in the space near the body require the implementation of particularly appropriate and precise behaviors: be it for dealing with an imminent threat or for attaining an object of interest.

PPS has been thought to have different roles; a protective role by implementing a safety margin around the body, enabling the preparation of defensive behaviors (Dosey & Meisels, 1969; Graziano & Cooke, 2006); a communicative role by determining the quality and quantity of sensory information exchanged during social interaction (Aiello, 1987; Hall,

1966) and a role in action execution (Brozzoli et al., 2012). The different roles that are suggested for peri-personal space are not totally independent and could be gathered around a more general role in the implementation of appropriate and precise approach and avoidance behaviors for an optimal physical interaction with the external world.

4.2. Affective events in the space around us

4.2.1. Processing of affective events located at close distances from the body

A few studies have explored the processing of affective events when they are located at close distances as opposed to farther distances from the body. A neuroimaging study has revealed an increased activity in different brain areas in response to aversive events located near the hand. In this study, participants saw a painful object (a syringe) or a non-painful object (a cotton bud) touching a realistic rubber hand placed either on the top of their real hand or at an incongruent spatial location with their real hand. Participants did not receive any tactile stimulation. Provided that the rubber hand was placed in a congruent spatial location with participants' real hand, cerebral activity, as measured with fMRI, was increased within the PCC, the mid cingular cortex and the anterior insula in response to painful versus non-painful stimuli (Lloyd, Morrison, & Roberts, 2006). These findings suggest that the presence of an aversive stimulus at close distances from the body increases activation in brain areas known to be involved in the preparation of appropriate motor responses and in affective processing.

In a study by Schiffenbauer and Shiavo, the effect on affective experience of a close versus far distance from an unpleasant or pleasant individual was investigated. Participants in this study had to perform a problem-solving task whilst another individual was sitting and observing from either a close or far distance from them. After the task, the observer, who actually was an experimenter, reported that he found that participants' strategy in solving the problem to be either stupid or smart. Then, participants had to rate how much they liked the other individual using scales. The experimenter was less liked when he commented that the participant's strategy was stupid than when he said that participants' strategy was smart. Furthermore, the spatial location of the experimenter in relation to the participants modulated the liking responses. In the negative condition, the experimenter was even more disliked if he was sitting close to the participant. Similarly, in the positive condition, the experimenter was more liked if he was sitting close as opposed to far from the participants. This suggests that both negative and positive affective events induce amplified affective experiences when located at close distances from the body (Schiffenbauer & Steven Schiavo, 1976).

The study of Williams and Bargh (2008) can be connected to the two previous studies even though they tested the effect of distance without reference to the body. They investigated the effect of spatial cues on affective experiences in response to violent media. First, their participants were primed with either spatial closeness or spatial distance by means of points located either close or far from each other on a Cartesian plane coordinate system before reading a violent excerpt from a book. Then, using a questionnaire, participants reported their feelings. They observed that participants primed with spatial closeness reported more negative feelings in response to the book excerpt than participants primed with spatial distance; this suggests that perception of spatial proximity between two objects amplifies the affective experience induced by stimuli evocating negative affect (Williams & Bargh, 2008).

* * *

Even though research in the processing of affective events located within close distances is sparse for the moment, the studies reported here suggest a modulation of affective processes according to the spatial location of the emotional event, i.e. in the within close distances or at farther distances.

4.2.2. Modulation of peri-personal space boundaries in the presence of affective events

Modulation of peri-personal space boundaries

Research has revealed that the boundaries of PPS are not rigid but rather highly flexible. The size of the area, which is represented as a zone that is near the body, can vary for a same individual according to situational factors. For example, the use of tools can modify the size of PPS (see Figure 4.3 for an example). Tools can allow for the reach of objects that would be otherwise out of range, and thus extend the space of possible physical interaction. Many studies have demonstrated that tool-use promotes a remapping of far distances within the boundaries of the PPS in both humans and non-human primates (e.g. Bassolino, Serino, Ubaldi, & Làdavas, 2010; Iriki, Tanaka, & Iwamura, 1996; Ladavas & Serino, 2008; Maravita & Iriki, 2004; Serino, Bassolino, Farnè, & Làdavas, 2007). Conversely, reducing the spatial range of possible physical action has been shown to shrink PPS (Lourenco & Longo, 2009). Besides this action-dependent plasticity of PPS; A study manipulating the quantity of sensory information exchanged during social interaction has revealed that the adopted interpersonal distance (i.e. PPS), is larger when auditory information from the interlocutor is not available (Lloyd, Coates, Knopp, Oram, & Rowbotham, 2009).

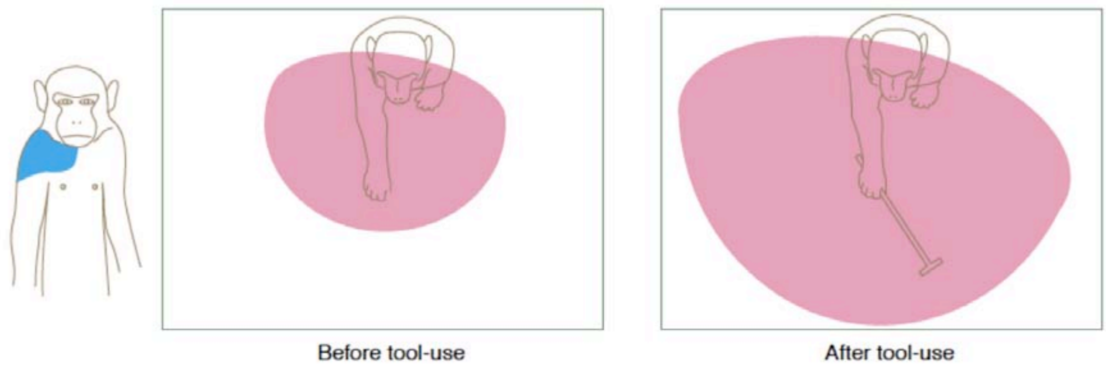

Figure 4.3. Extension of peri-personal space through tool-use in the monkey (from Maravita & Iriki, 2004).
Tactile (blue) and visual (pink) receptive fields of a neuron in the intraparietal cortex. Immediately after tool-use, the visual receptive field of the neuron is extended and includes farther distances from the tactile receptive field on the body.

Modulation of peri-personal space boundaries in the presence of affective events

Given the close relationship of PPS with defensive (Graziano & Cooke, 2006) and approach (Brozzoli et al., 2012) related behaviors, the influence of the presence of affective stimuli on the size of PPS has emerged as a topic of interest. Several studies have addressed this issue with different experimental paradigms.

With stop-distance tasks, wherein participants are instructed to stop the approach of another individual (see Figure 4.4), the affective state expressed by the person approaching has been shown to modulate the interpersonal distance of comfort. For example, a greater distance was maintained with hostile individuals who insulted participants, suggesting an extension of PPS in the presence of unpleasant people (O'Neal, Brunault, Marquis, & Carifio, 1979). Conversely, experimental paradigms consisting of an unobtrusive observation of the distances used during social interaction revealed that participants used closer distances during interaction with individuals evaluated as friendly, attractive or cooperative, suggesting a contraction of PPS in the presence of pleasant people ((Byrne, Erwin, & Lamberth, 1970; Gifford, 1982; see also Tedesco & Fromme, 1974 as quoted by Aiello, 1987).

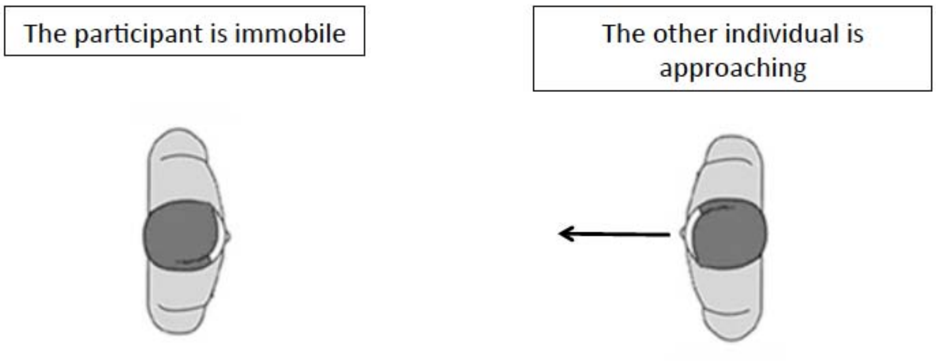

Figure 4.4. Stop-approach task to assess peri-personal space size.
The participant is approached by another person (generally an experimenter). The participant's task is to stop the approach as soon as he/she experiences discomfort. The distance, at which the approaching person is stopped, is considered as an indicator of the location of participant's peri-personal space boundaries.

In another type of study, PPS size was assessed in empirical situations wherein both participants and affective stimuli were immobile (see Figure 4.5). In their study, Valdés-Conroy and colleagues presented digitalized objects on a horizontal screen on a table at different distances from their participants; they then instructed the participants to indicate whether or not they thought that the objects were close enough to be reached. The objects used were either evaluated as positive (chocolates, a diamond, a ring, a hamburger), neutral (a knitting ball, a paper-clip, buttons, brush) or negative (a rotten orange, cigarette-buds, flies, excrements) and the perceived reachability of these objects was taken as an indication of the representation of PPS boundaries. The findings point to an extension of PPS in the presence of appetitive objects and a reduction in the presence of aversive objects (Valdés-Conroy, Román, Hinojosa, & Shorkey, 2012). Using a similar paradigm, the contraction of PPS in the presence of aversive objects was also observed with a syringe, scissors, a box cutter and a cork screw presented with the dangerous side pointing towards participants' body (Coello, Bourgeois, & Iachini, 2012).

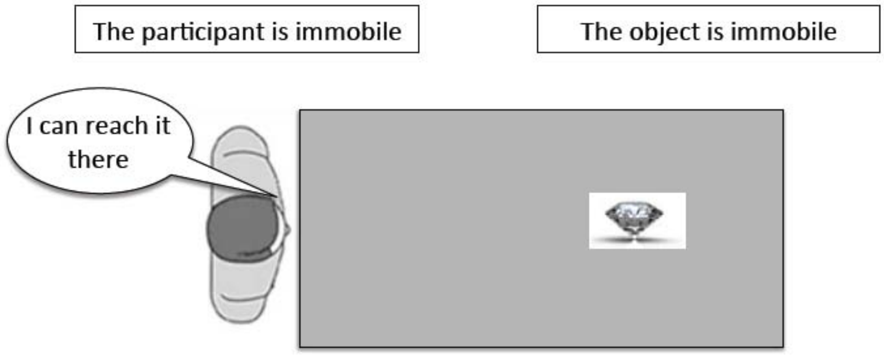

Figure 4.5. Subjective evaluation of reachability to assess peri-personal space size (insprired by Valdés-Conroy, 2012)
The participant is sited and objects are presented on a table in front of him and at different distances from him. The participant's task is to indicate whether he thinks that the objects are close enough for him to reach. The distances, at which the objects are evaluated as reachable, are considered as located within peri-personal space boundaries.

In another study, Teneggi et al. replicated the extension of PPS in the presence of a positive stimulus using also an experimental situation where both participants and stimuli were immobile. However, they used a method different from the subjective evaluation of reachability to measure PPS size (see Figure 4.6). The method that was used is based on the multisensory quality of PPS representation and assumes that the spatial location at which surrounding sensory events start to be integrated with sensory events located on the body reflects the boundaries of PPS (Canzoneri, Magosso, & Serino, 2012). They observed that their participants' PPS was extended in the presence of another individual after having had a satisfying social interaction between them (Teneggi, Canzoneri, Di Pellegrino, & Serino, 2013), placing the other inside PPS boundaries.

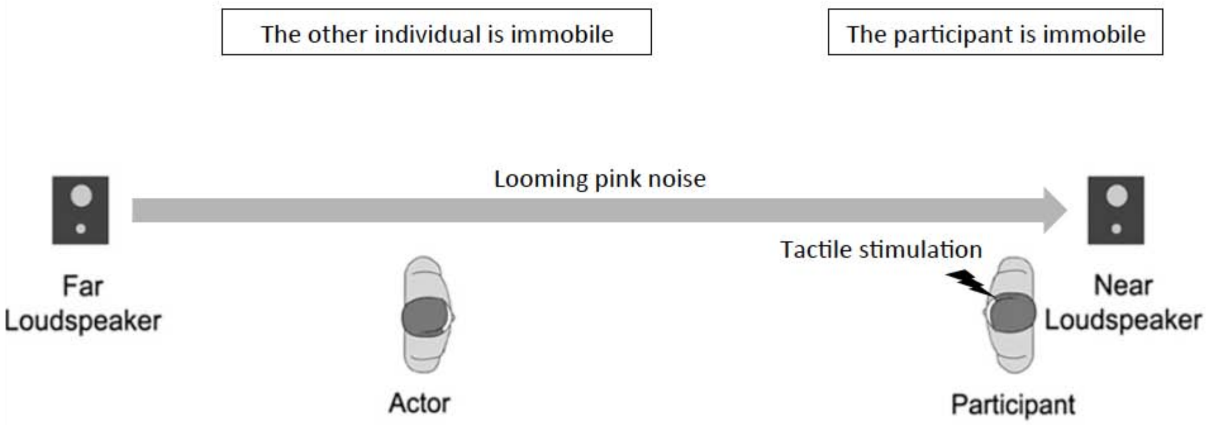

Figure 4.6. Auditory-tactile task to assess peri-personal space size (adapted from Teneggi et al., 2013)
The participant is standing and facing another individual. A sound consisting of pink noise is looming toward him. His task is a speeded detection of a tactile stimuli delivered on his cheek. The tactile stimuli are delivered when the sound source is located at different distances from the participant. When the sound source is located within peri-personal space, there is an auditory-tactile interaction and the behavioral response to tactile stimulation is faster.The distance between the participant and the sound source when the sound starts to speed up tactile detection is considered as the location of participant's peri-personal space boundaries.

Perception of looming affective events

Several other studies provide findings that may be linked to an influence of the presence of affective stimuli on PPS size. Two studies examined the perception of looming visual stimuli as a function of their emotional significance. They employed paradigms wherein participants were immobilized whereas the affective stimuli were looming towards them. The looming movement was simulated by expanding pictures of affective stimuli. The task was a time-to-collision estimation task. Participants had to indicate the moment at which they thought the visual stimulus would collide with them. Brendel and colleagues presented looming pictures depicting aversive or neutral scenes to their participants. They found that participants underestimated the time-to-collision with threatening stimuli as compared to neutral ones (Brendel, DeLucia, Hecht, Stacy, & Larsen, 2012). Similar conclusions were drawn from a study using pictures of snakes and spiders as aversive stimuli and pictures of butterflies and rabbit as non-aversive stimuli (Vagnoni, Lourenco, & Longo, 2012). Furthermore, this study indicated that the more participants feared the aversive stimuli the more they underestimated the time-to-contact. These studies did not investigate PPS size. However, the collision with the stimuli that participants had to predict may be related to the entry of the stimuli within PPS boundaries. The underestimated time-to-collision with aversive events could then be linked to an extension of PPS boundaries.

Affective state and peri-personal space boundaries

Three other studies may be related to the investigation of the effect of affective stimuli on PPS size. Although these studies did investigate PPS size, they did not examine the role of the presence of affective stimuli on PPS size. Instead, they explored the effect of participants' affective state on the size of their PPS. The affective state induced in these studies can be paralleled to the affective state induced by the presence of an affective stimulus.

Dosey and Meisels induced a negative affective state in their participants using performance evaluation. They observed that participants stressed by the evaluation adopted larger interpersonal distances in a stop-distance task wherein they had to approach another individual (Dosey & Meisels, 1969). In another study, Tajadura-Jiménez and colleagues found that participants, who were in a positive affective state that was induced by music, tolerated closer interpersonal distance in a stop-distance task where they were approached by another person (Tajadura-Jiménez, Pantelidou, Rebacz, Västfjäll, & Tsakiris, 2011).

Instead of interpersonal distances, Gagnon and colleagues used judgments of reachability to assess PPS size according to the affective state of participants. Negative affective states were induced by instructing participants to recall a situation in which they experienced fear and to write about it. In a control condition, participants had to describe their morning routine. Gagnon and colleagues showed that reachability judgments of an event localized *via* auditory cues was influenced by the affective state of the participants: when in a negative affective state, their PPS was smaller (Gagnon, Geuss, & Stefanucci, 2013).

* * *

The body of work described here provides strong evidence that the spatial area represented as peri-personal has highly flexible boundaries, which can be influenced by the presence of affective events in the environment. However, the findings are not in agreement about the effect of affective stimuli valence (negative or positive) on the direction of PPS boundaries shift, i.e. leading to an extension or a contraction of PPS. Nevertheless, it is possible that this variability of results is explained by the diversity of experimental paradigms and tasks that have been adopted in the different studies. Whether participants are immobilized or not as well as whether the affective events is mobile or not in the experimental setups may lead to a preferential engagement of approach or defensive behaviors.

The localization of affective events in the environment is of critical importance for implementing appropriate behaviors and interactions. Research has demonstrated that space is not a unitary construct. Events located in the area close to our bodies are represented differently from those located at farther distances from the body. Even if sparse, there is some evidence suggesting that affective processing is modulated depending on the distance of affective events' from the body.

The specific representation of the area surrounding the body, also called peri-personal space, has been suggested to have a role in the implementation of defensive and approach behaviors. The size of peri-personal space is flexible and may be modified in the presence of affective events in the environment.

EXPERIMENTAL CONTRIBUTIONS

The primary goal of this research work was to investigate the influence of multisensory stimuli on conscious emotional experience. Given that natural situations involve a spatial dimension, this work also explored the relationship between space and affect.

Three studies were conducted:

- **Study A** explored the influence of multisensory stimulation on emotional experience with virtual reality techniques. The effect of the auditory-visual presentation of aversive stimuli on negative emotional experience was investigated with the hypothesis that emotional experience would be amplified by multisensory stimulations.

- **Study B** explored the influence of excessive fear on space representation. The effect of cynophobic fear (dog fear) on the size of peri-personal space was tested with the hypothesis that cynophobic-based anxiety would modulate the extent of peri-personal space.

- **Study C** explored the influence of multisensory stimulation on emotional experience as a function of the distance to the aversive object with virtual reality techniques. The effect of auditory-visual aversive stimuli on negative feelings was investigated; as the stimulus was located close or far away, this included the hypothesis that the distance to the aversive object modulates the influence of multisensory stimulation on emotional experience.

5. GENERAL METHODOLOGY

5.1. Stimuli

Given that ecological validity and semantic congruency are important factors in multisensory processes (De Gelder & Bertelson, 2003; Laurienti et al., 2004), natural multisensory events were used to induce negative feelings during the experimentation. We chose to circumvent facial and vocal stimuli, which are highly specific affective events, in order to broaden the investigation to other natural multisensory events, such as a dog or a crowd. Dogs and crowds naturally are multisensory stimuli since they can be perceived *via* both audition and vision. These stimuli, however, are not necessarily evaluated as fearsome and do not necessarily induce a negative affective state. Dogs are often sought as companions and joining a crowd of people is often desirable in order to celebrate events such as a new year or a win after a sports competition. Nevertheless, dogs and crowds remain of evolutionary emotional relevance. Dogs and crowds can threaten life. Whereas dogs and crowds are common fear-relevant stimulations, individuals who are phobic of dogs or crowds consider them as genuinely fearsome. Phobias are anxiety disorders, which are characterized by an unreasonable fear feeling in response to a feared object or situation. Given the objective threat, the intensity of phobic patients' fear is unreasonable. Specific phobias such as cynophobia (dog phobia) and crowdphobia are thought to be more so fear-related rather than anxiety-related disorders because it is possible to explain them with one central problem in the responsiveness of the defensive system (Lewis et al., 2008). In order to ensure that the experiments induce negative feelings in the participants, we recruited individuals who were sensitive to cynophobia and crowdphobia, in accordance with the experiment. They were all non-pathological individuals who were specifically fearful of the natural multisensory stimuli that we presented to them.

5.2. Participants

For the three studies presented below, we selected healthy participants on the basis of their responses on questionnaires evaluating stimulus-based anxiety. In Studies A and B, participants were selected among a sample of individuals who completed a questionnaire assessing cynophobic-based anxiety (the fear of dogs). This dog phobia questionnaire (Viaud-

Delmon et al., 2008) consists of two sections. The first section asked four yes/no questions about reactions to dogs: "Do you fear dogs more than other people do?", "Do you endure the presence of dogs with anxiety?", "Are you afraid of a specific dog breed and if yes, which one?", "Does the size of the dog have an effect on your fear?". 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 the questionnaire is 0, with a maximum of 42. The French version of the dog phobia questionnaire can be found in the Annex section. In Study C, participants were selected on the basis of their score on a questionnaire assessing crowdphobic-based anxiety (the fear of crowds). The French version of the crowdphobia questionnaire can be found in the Annex section.

All of our participants took part in a diagnostic interview with a clinical psychologist based on the Mini International Neuropsychiatric Interview. This interview was conducted to ascertain that no participant met criteria for pathological anxiety disorders, and thus avoid biases in the investigation of negative emotional experience.

5.3. Methodology specific to the studies in virtual reality

Stimuli

The virtual stimuli (dog and crowd stimuli), used in the virtual reality studies, were chosen because they can convey affective information *via* both auditory and visual pathways. They were presented embedded in auditory-visual virtual environment depicting natural matching-context. The visual component of stimuli was rendered in 3D with visual stereoscopy. The auditory component of stimuli was also rendered in 3D through binaural rendering. Hence, virtual stimuli could be localized on the basis of both visual and auditory cues. The sensory presentation of the stimuli was manipulated so as to display credible natural situations.

Measures of affective responses

We measured the intensity of negative emotional experience in response to virtual stimuli with Subjective Units of Distress (SUD: Wolpe, 1973; see chapter 4 for more details). To make self-reporting easier for the participants, we instructed them to use a scale from 0 to 10 and then transferred the measures on a scale from 0 to 100.

Behavioral Avoidance/Assessment Test (BAT)

The BAT is a widely used technique in clinical psychology to assess the level of fear of phobic patients and to evaluate the treatment success by comparing the results of the test before and after therapy (see chapter 4 for more details). This BAT is also used in the field of virtual-reality based therapy for anxiety disorders. The subject is immersed in a virtual environment and is tasked with approaching the fear-object, which is in that case a virtual object (see Mühlberger, Sperber, Wieser, & Pauli, 2008 for an example of BAT in virtual reality). In Studies A and C, we used BATs in virtual reality to evaluate the behavioral component of participants' stimulus-based anxiety (dog fear and crowd fear) and to assess an eventual habituation phenomenon after the experiment.

Questionnaires related to the immersion in virtual reality

A 22-item cybersickness scale was used to assess participants' level of discomfort linked to the use of virtual reality setups (Viaud-Delmon et al, 2000). This questionnaire comprises a list of symptoms and sensations associated with autonomic arousal (nausea, sweating, heart pounding, etc.), vestibular symptoms (dizziness, fainting, etc.), respiratory symptoms (feeling short of breath, etc.) and can also be used to estimate signs of somatisation (tendency to complain of a large number of diverse symptoms). Items are rated on a scale from 0 to 4 (absent, weak, moderate, strong). The minimal score on this questionnaire is 0, with a maximum of 88.

We evaluated participants' experience of presence in the virtual environments with the presence questionnaire from the I-group (Schubert et al 2001). This questionnaire is composed of 14 items related to their mental, perceptive and emotional state evoked by the fact that they were isolated from the outside world with only virtual information. Each item is rated on a scale from 0 to 7. The minimal score on this questionnaire is 0, with a maximum of 84. The French versions of these questionnaires can be found in the Annex section.

6. AUDITORY-VISUAL STIMULATION AND SENSITIVITY TO DOG PHOBIA

Exploring the effect of multisensory affective stimuli on emotional experience with virtual reality

6.1. Description and main findings of the study

As described in the general introduction, the number of studies exploring the affective processing of multisensory stimuli has been growing for the past 20 years. These studies have mostly concentrated on the first steps of affective processing in highly controlled paradigms involving stimuli such as pictures, video and sound recordings. Study A attempts to contribute to this field by investigating the effect of the multisensory presentation of stimuli on the conscious emotional experience in ecological and interactive situations.

More specifically, we investigated whether the auditory-visual presentation of aversive stimuli modulates the conscious experience of fear. Subjective measures of fear (i.e. SUD) were collected in response to auditory-only, visual-only and auditory-visual dog stimuli. We used the unique advantages of virtual reality techniques to present the dog stimuli embedded in a natural context and to control their display in terms of sensory presentation. We recruited healthy participants to take part in the study. We constituted a group of 12 individuals sensitive to cynophobia (dog-fearful group) for whom dogs are genuinely aversive and a control group of 10 individuals non-sensitive to cynophobia for whom dogs are not aversive but still fear-relevant. Both groups of participants encountered the dog stimuli during a navigation task in two virtual environments (a garden virtual scene and a hangar virtual scene).

Results showed that the sensory presentation of the aversive stimuli significantly affected the subjective ratings of fear. Individuals sensitive to cynophobia as well as individuals non-sensitive to cynophobia experienced more fear when encountering bimodal dog stimuli as compared to unimodal dog stimuli. These results suggest that the multisensory presentation of stimuli amplifies the experience of emotion. Given that the fear in response to auditory-visual dog stimuli was significantly higher than the sum of the fear in response to auditory and visual dog stimuli, it is possible that this phenomenon is linked to cross-modal potentiation.

6.2. Paper A

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: $M = 37.09$, $\text{SD} = 13.78$) 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.

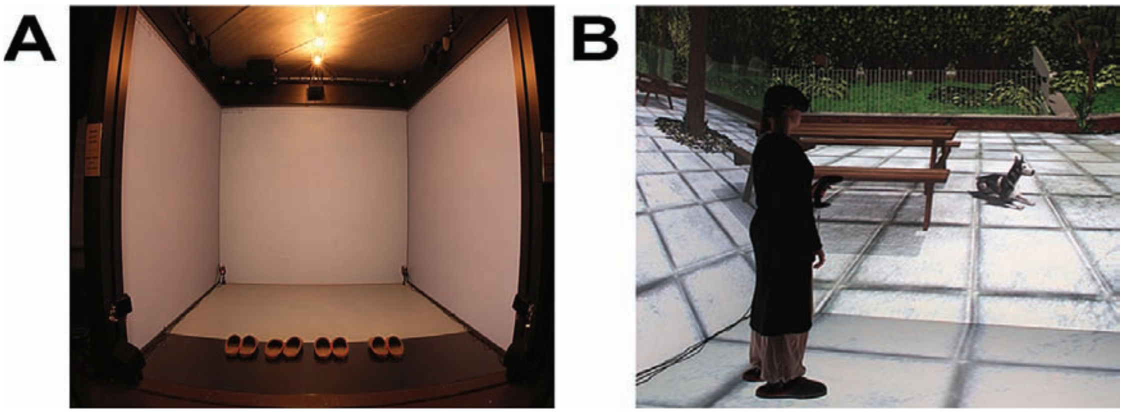

Figure 1. (A) Picture of the iSpace setup used in the study. (B) A participant, equipped with polarized glasses, headphones and a wireless joystick, standing within the iSpace during immersion in an auditory–visual VE. This figure is published in colour in the online version.

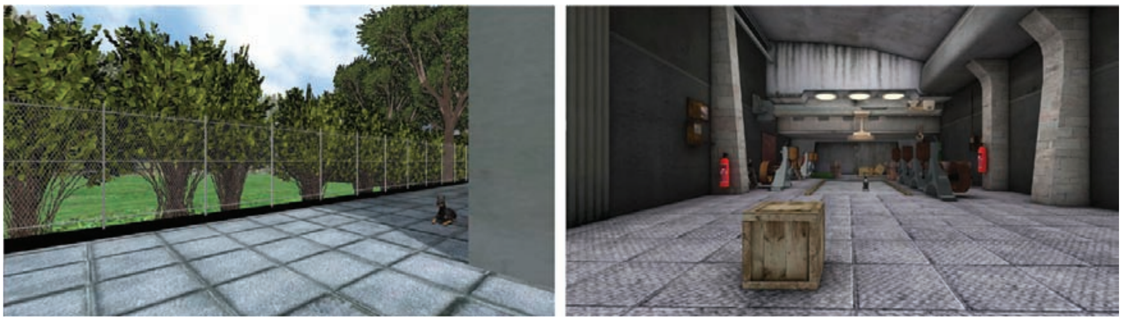

Figure 2. Pictures of the auditory–visual VEs used to measure the participants' fear when encountering virtual dogs. On the left, the outdoor garden scene and on the right, the indoor hangar scene. This figure is published in colour in the online version.

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: $M = 31.71$, $\text{SD} = 11.40$) completed this questionnaire. A mean dog fear score ($M = 10.63$, $\text{SD} = 8.55$) 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 online. 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 ($n_{\text{NoFear}} = 10$; $n_{\text{DogFear}} = 9$), 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 ($n_{\text{NoFear}} = 9$; $n_{\text{DogFear}} = 9$). We compared cybersickness and presence scores between groups ($n_{\text{NoFear}} = 10$; $n_{\text{DogFear}} = 9$).

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

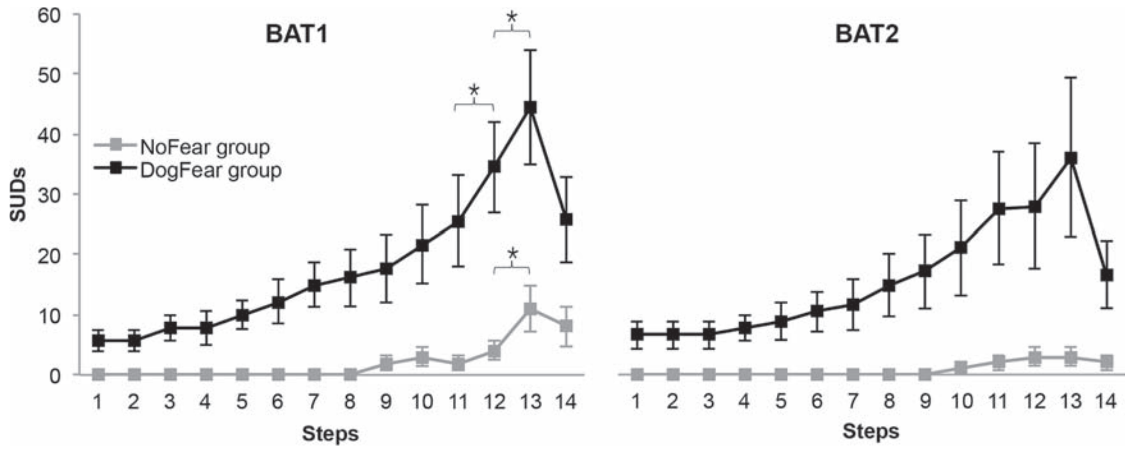

Figure 4. Mean reported fear (mean SUDs ± SEM) of the NoFear group (grey squares) and the DogFear group (black squares) at each of the 14 steps during BATs. The responses collected during BAT1 are presented on the left and the responses collected during BAT2 are presented on the right. In the NoFear 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 ± 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

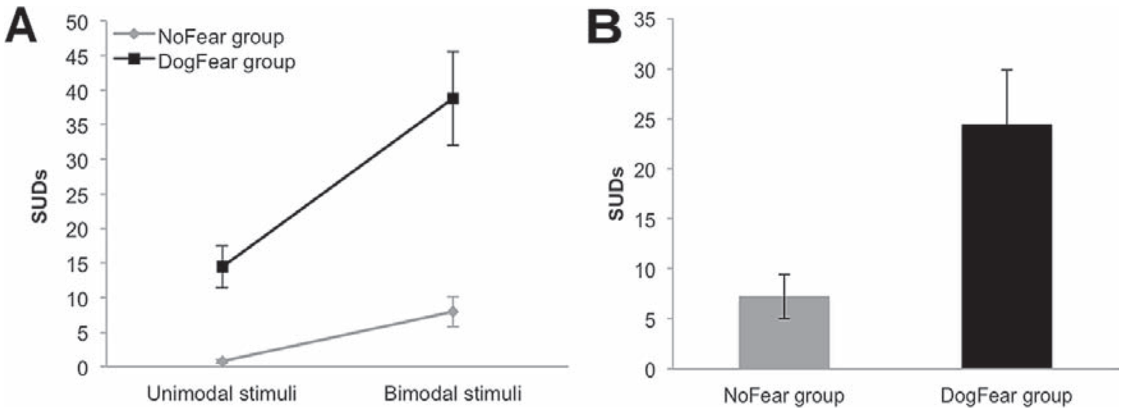

Figure 5. (A) Mean reported fear (mean SUDs ± SEM) of the NoFear (grey diamonds) and DogFear group (black squares) in the auditory–visual VEs according to the sensory modality in which the dogs were presented. The SUDs reported in response to the auditory static, the visual static, the auditory moving 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 ($T = 18.50$, $p = 0.006$; $M_{\text{Bimodal}} = 22.59$, $SD_{\text{Bimodal}} = 21.35$; $M_{\text{Sum Unimodal}} = 14.96$, $SD_{\text{Sum Unimodal}} = 19.25$). 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 ($T = 25.00$, $p = 0.272$; $M_{\text{Unimodal/Outdoor}} = 6.32$, $SD_{\text{Unimodal/Outdoor}} = 8.91$; $M_{\text{Unimodal/Indoor}} = 8.42$, $SD_{\text{Unimodal/Indoor}} = 10.55$).

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 ($T = 24.00$, $p = 0.239$, $M_{\text{Outdoor}} = 0.088$, $SD_{\text{Outdoor}} = 0.111$; $M_{\text{Indoor}} = 0.027$, $SD_{\text{Indoor}} = 0.140$; $N = 12$). The nSCL was lower during unimodal stimulations ($M_{\text{unimodal}} = 0.038$, $SD_{\text{unimodal}} = 0.117$, $N = 12$) compared to bimodal stimulations ($M_{\text{bimodal}} = 0.077$, $SD_{\text{bimodal}} = 0.099$, $N = 12$) 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 ($T = 0.00$, $p = 0.008$; $M_{\text{DogFear/barking}} = 15.00$, $SD_{\text{DogFear/barking}} = 10.31$; $M_{\text{DogFear/growling}} = 38.33$, $SD_{\text{DogFear/growling}} = 21.65$). In the NoFear group, there was no significant difference between the SUDs in response to the barking and to the growling sound ($T = 3.00$, $p = 0.465$; $M_{\text{NoFear/barking}} = 9.44$, $SD_{\text{NoFear/barking}} = 21.13$; $M_{\text{NoFear/growling}} = 12.22$, $SD_{\text{NoFear/growling}} = 16.60$).

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.

7. AUDITORY-TACTILE STIMULATION AND SENSITIVITY TO DOG PHOBIA

Exploring the influence of excessive fear on space representation

7.1. Description and main findings of the study

It has been proposed that peri-personal space (PPS), i.e. the space surrounding our body, is involved in implementing a safety margin around the body allowing for the preparation of defensive behaviors against unwanted intrusions. Given that PPS boundaries are flexible, it is possible that anxiety influences PPS size. Studies investigating this question have found contradicting results and the influence of anxiety on PPS size still is an open research topic.

We studied the effect of cynophobic-based anxiety on PPS boundaries in the presence of a dog stimulus. For this study, we recruited a non-clinical sample of individuals on the basis of their sensitivity to cynophobia. We constituted a group of 15 healthy individuals sensitive to cynophobia (dog-fearful group) and a group of 15 healthy individuals non-sensitive to cynophobia (non-fearful group). We used an audiotactile task to dynamically measure the extent of participants' PPS when in the presence of a dog auditory stimulus and when in the presence of a sheep auditory stimulus. The virtual sound sources of the dog and the sheep stimuli loomed towards participants from the rear hemi-field.

Results showed that PPS size in the presence of a sheep stimulus was similar in the dog-fearful and the non-fearful groups. In the presence of a dog stimulus, the PPS boundaries of participants with excessive fear of dogs extended. This effect of the dog stimulus on PPS boundaries was not observed in the non-fearful group. These findings demonstrate that PPS size is adaptively modulated by sensitivity to cynophobia and suggest that anxiety tailors PPS boundaries when exposed to fear-relevant features.

7.2. Paper B

frontiers in
PSYCHIATRY

ORIGINAL RESEARCH ARTICLE
published: 03 September 2014
doi: 10.3389/fpsyt.2014.00122

Cynophobic fear adaptively extends peri-personal space

Marine Taffou[1,2,3,4,5,6,7]* **and Isabelle Viaud-Delmon**[1,2,3]

[1] CNRS, UMR 9912, STMS, Paris, France
[2] IRCAM, UMR 9912, STMS, Paris, France
[3] Sorbonne Universités, UPMC Univ Paris 06, UMR 9912, STMS, Paris, France
[4] Inserm, U 1127, Paris, France
[5] CNRS, UMR 7225, Paris, France
[6] Sorbonne Universités, UPMC Univ Paris 06, UMR S 1127, Paris, France
[7] Institut du Cerveau et de la Moelle épinière, ICM, Social and Affective Neuroscience (SAN) Laboratory, Paris, France

Edited by:
Pablo Billeke, Universidad del Desarrollo, Chile

Reviewed by:
Matthew R. Longo, Birkbeck University of London, UK
Elisa Canzoneri, EPFL, Switzerland

***Correspondence:**
Marine Taffou, UMR 9912 CNRS IRCAM UPMC, Institut de Recherche et Coordination Acoustique/Musique, 1 Place Igor Stravinsky, Paris 75004, France
e-mail: marine.taffou@ircam.fr

Peri-personal space (PPS) is defined as the space immediately surrounding our bodies, which is critical in the adaptation of our social behavior. As a space of interaction with the external world, PPS is involved in the control of motor action as well as in the protection of the body. The boundaries of this PPS are known to be flexible but so far, little is known about how PPS boundaries are influenced by unreasonable fear. We hypothesized that unreasonable fear extends the neural representation of the multisensory space immediately surrounding the body in the presence of a feared object, with the aim of expanding the space of protection around the body. To test this hypothesis, we explored the impact of unreasonable fear on the size of PPS in two groups of non-clinical participants: dog-fearful and non-fearful participants. The sensitivity to cynophobia was assessed with a questionnaire. We measured participants' PPS extent in the presence of threatening (dog growling) and non-threatening (sheep bleating) auditory stimuli. The sound stimuli were processed through binaural rendering so that the virtual sound sources were looming toward participants from their rear hemi-field. We found that, when in the presence of the auditory dog stimulus, the PPS of dog-fearful participants is larger than that of non-fearful participants. Our results demonstrate that PPS size is adaptively modulated by cynophobia and suggest that anxiety tailors PPS boundaries when exposed to fear-relevant features. Anxiety, with the exception of social phobia, has rarely been studied as a disorder of social interaction. These findings could help develop new treatment strategies for anxious disorders by involving the link between space and interpersonal interaction in the approach of the disorder.

Keywords: emotion, anxiety, cynophobia, auditory–tactile integration, multisensory integration, spatial audition, 3D sound, looming sound

INTRODUCTION

Peri-personal space (PPS) is defined as the space immediately surrounding our bodies (1), through which interaction with the external world occurs. PPS is opposed to the more distant, extra-personal space. Studies on both monkeys and humans have supported this distinction by showing that stimuli within PPS are represented distinctly in the brain from stimuli within extra-personal space (2). In the field of social psychology, this space near the body is referred to as "personal space" and has been described as an area with invisible boundaries that individuals actively maintain around themselves, into which the intrusion of unwanted stimulation causes discomfort (3, 4). It has been proposed that one of the roles of PPS is to implement a safety margin, which allows for the preparation and coordination of defensive behaviors against unwanted intrusions (2, 5).

Recent studies have brought evidence that the boundaries of PPS are flexible. For example, PPS can be extended through tool-use (6–8), by satisfying social interaction with others allowing integrating them to one's PPS (9) or by depriving individuals of auditory cues from the external world (10). PPS can also be shrunk by increasing the effort needed to perform a hand movement with wrist weights (11) or by listening to positive emotion-inducing music through headphones leading to a better tolerance of others' proximity (12).

In the present study, we investigated whether PPS size is influenced by anxiety. We hypothesized that the disproportionate experience of fear observed in some anxious disorders may be linked to the introduction of the fear-object in the boundaries of the individual's exaggerated PPS. We explored the impact of cynophobic-based anxiety, i.e., the excessive fear of dogs on the size of PPS in two groups of non-clinical participants: dog-fearful and non-fearful participants. We recruited two groups of individuals – individuals sensitive to cynophobia [dog-fearful (DF) group] and individuals non-sensitive to cynophobia [non-fearful (NF) group] – and measured the extent of their PPS in the presence of threatening (dog growling) and non-threatening (sheep bleating) auditory stimuli looming from the rear hemi-field. Participants performed a tactile detection task with their

Table 1 | Participants' characteristics.

Variable	All participants	NF group	DF group
Number of individuals	$N = 30$	$n_{NF} = 15$	$n_{DF} = 15$
Number of female[a]	24	10	14
Age ($M \pm SD$)[a]	25.60 ± 7.73	26.93 ± 9.15	24.27 ± 6.03
95% Confidence interval	(22.71; 28.49)	(21.87; 32.00)	(20.93; 27.61)
Trait anxiety score ($M \pm SD$)[b]	40.53 ± 9.79	35.53 ± 9.13	45.53 ± 7.84
95% confidence interval	(36.88; 44.19)	(30.48; 40.59)	(41.19; 49.87)
Dog fear score ($M \pm SD$)[c]	15.33 ± 13.66	2.40 ± 1.64	28.27 ± 5.02
95% Confidence interval	(10.23; 20.43)	(1.49; 3.31)	(25.49; 31.05)
Range	(0.00; 36.00)	(0.00; 5.00)	(21.00; 36.00)

[a] *Both groups were similar in terms of ratio of female [χ^2 test with Yates correction: $\chi^2_{(1)} = 1.88$, $p = 0.171$] and age [T-test: $t_{(28)} = -0.94$, $p = 0.354$].*

[b] *The trait anxiety score was significantly different between groups [T-test: $t_{(28)} = 3.22$, $p = 0.003$, $d = 1.18$].*

[c] *The variance of dog fear scores was different between groups [$F_{(15,15)} = 9.39$, $p = 0.0002$], hence a non-parametric test was conducted. The dog fear score was significantly different between groups (Mann–Whitney U-test: $U = 0.00$, $p < 0.001$).*

left hand while the task-irrelevant sounds were looming toward them from the rear hemi-field. The measure of rear PPS boundaries with this audiotactile task is particularly appropriate since the auditory component of looming stimuli is especially relevant in the rear hemi-field, where the visual monitoring is not possible.

MATERIALS AND METHODS

PARTICIPANTS

Participants were selected on the basis of their scores on a questionnaire exploring the fear of dogs (13). The minimal score on this dog fear questionnaire is 0, with a maximum of 42. Four hundred eighteen individuals (236 females; age: $M = 28.87$, $SD = 10.44$) completed this questionnaire. A mean dog fear score ($M = 11.67$, $SD = 9.19$) as well as a median dog fear score (Median = 8) were obtained from the questionnaire results, which served as a basis to select participants for the current experiment. Thirty healthy individuals (see details in **Table 1**) with normal audition and touch participated in the study. All of them were right-handed. None of them had a history of psychiatric disorders, neurological disorders or was currently undergoing medical treatment. Fifteen individuals had a low dog fear score (score <20th centile) and thus composed the NF group. The remaining 15 individuals had high dog fear scores (score >80th centile) and composed the DF group. We also used the State Trait Anxiety Inventory (STAI) (14) to measure anxiety levels. Participants completed the trait version several weeks before the experiment. All participants provided written informed consent prior to the experiment, which was approved by the Health Research Ethics Committee (CERES) of Paris Descartes University. Participants were paid 10 €/h.

EXPERIMENTAL SETUP AND STIMULI

We used a modified version of Canzoneri et al.'s audiotactile interaction task (15). Participants were blindfolded and sat on a chair with their hands palms-down on a table. Both of their hands were aligned with their mid-sagittal plane. Head movements were minimized by means of a headrest.

Auditory stimuli were presented through Sennheiser HD650 headphones. Auditory stimuli were two different (threatening and non-threatening) complex sounds (32 bits, 44100 Hz digitization). The threatening auditory stimulus was dog growling and the non-threatening one was sheep bleating. They were modified using audio editing software (Audacity software)[1] to be continuous 3000 ms sounds and to be similar in terms of temporal dynamic and amplitude. The auditory stimuli were then processed through binaural rendering using a non-individual head related transfer functions (HRTF) of the LISTEN HRTF database[2]. With this procedure, the virtual sound source location can be manipulated by rendering accurate auditory cues such as frequency spectrum, intensity, and inter-aural differences.

The tactile stimulus was a vibratory stimulus delivered by means of small loudspeaker on the palmar surface of the left index finger of participants. A sinusoid signal was displayed for 20 ms at 250 Hz. With these parameters, the vibration of the loudspeaker was perceivable, but the sound was inaudible. A PC running Presentation® software was used to control the presentation of the stimuli and to record the responses.

DESIGN AND PROCEDURE

First, participants were invited to take part in a 20 min long diagnostic interview with a clinical psychologist based on the Mini International Neuropsychiatric Interview. This interview was conducted to ascertain that no participant met criteria for pathological anxiety disorders. Following this interview, participants were invited to evaluate the valence and arousal of the sounds used in the main experiment. Afterwards they were asked to place their left index finger on the vibrator and to press a button with their right index finger each time a tactile stimulus was detected; this constituted the main experiment. At the end of the experiment, they were asked to again evaluate the valence and arousal of the sounds.

Main experiment

During the main experiment, an auditory stimulus was presented for 3000 ms for each trial. The sound source approached from the rear hemi-field, either from the right (135°) or from the left hemi-space (−135°), with a spatial location varying from 520 to 20 cm from the center of the participant's head. The auditory stimulus was preceded by 1000 ms of silence. A period of silence, with a duration varying between 2700 and 3300 ms, also occurred after the offset of the sound.

In 87.5% of the trials, a tactile stimulus was presented along with the auditory stimuli. The remaining 12.5% trials were catch trials with auditory stimulation only. Participants were instructed to ignore the auditory stimuli and to respond as quickly as possible

[1] http://audacity.sourceforge.net/
[2] http://recherche.ircam.fr/equipes/salles/listen/

to the tactile stimuli by pressing a button with their right index finger. They were asked to emphasize speed, but to refrain from anticipating. Reaction times (RTs) were measured.

Vibratory tactile stimuli were delivered at different delays starting from sound onset. With this procedure, the tactile stimuli were processed when the sound source was perceived at varying distances from participants' bodies. Given that a looming auditory stimulus speeds up the processing of a tactile stimulus as long as it is perceived near the body, i.e., within PPS (15), we considered the distance at which sounds boosted tactile RTs as a proxy of PPS boundaries.

Temporal delays for the tactile stimulus (see **Figure 1A**) were set as follows: T1 was a tactile stimulation administered simultaneously with the sound onset (corresponding to 1000 ms from the beginning of the trial); T2, at 750 ms from sound onset (at 1750 ms from trial beginning); T3, at 1500 ms from sound onset (at 2500 ms from trial beginning); T4, at 2250 ms from sound onset (at 3250 ms from trial beginning); and T5, at 3000 ms from sound onset (at 4000 ms from trial beginning). Thus, tactile stimulation occurred when the sound source was perceived at different locations with respect to the body, i.e., far from the body at low temporal delays and close to the body at high temporal delays (see **Figure 1B**). Moreover, in order to measure RTs in the unimodal tactile condition (without any sound), tactile stimulation was also delivered during the silent periods, preceding or following sound administration, namely at 350 ms (T_{before}) and at 4650 ms (T_{after}) after the beginning of the trial. The total test consisted of a random combination of eight target stimuli in each of the 28 conditions. The factors were: DELAY (seven levels: T_{before}, T1, T2, T3, T4, T5, T_{after}), HEMISPACE (two levels: left/right), and SOUND TYPE (two levels: threatening/non-threatening sound). There were a total of 224 trials with a tactile target, randomly intermingled with 32 catch trials. Trials were equally divided in 8 blocks of 32 trials, lasting about 4 min each. After each block, we verified that participants actually perceived the sounds as looming toward them from the rear hemi-field by directly asking them.

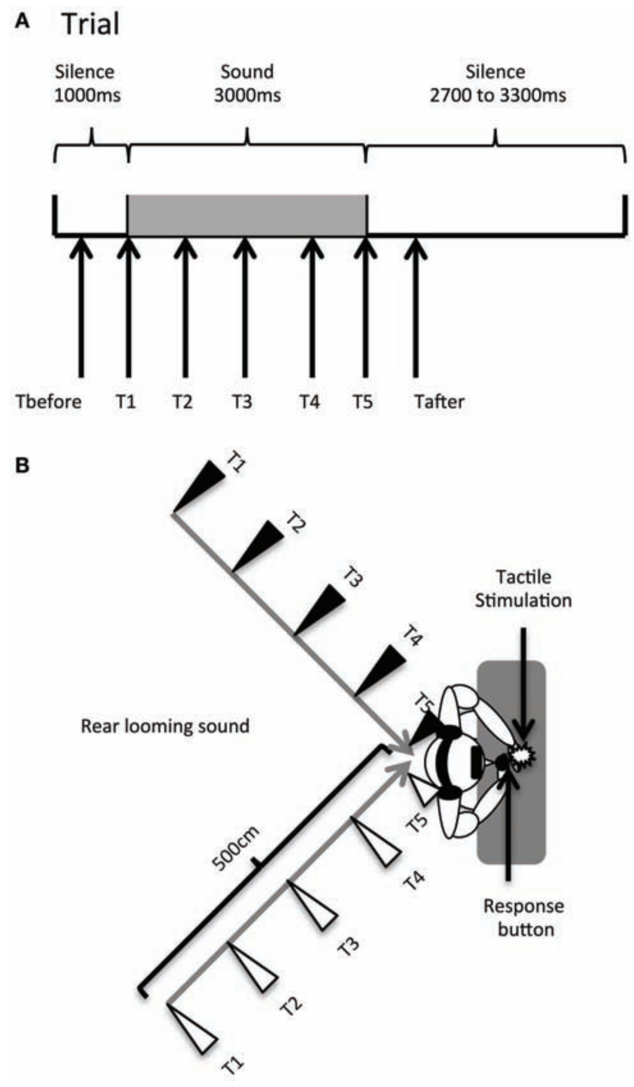

FIGURE 1 | Audiotactile test. (A). Description of a trial. **(B)** Experimental setup. Participants received a tactile stimulus at their hand while task-irrelevant sounds (threatening or non-threatening) approached them from the rear hemi-field, either in the left or the right hemi-space. When participants perceived the tactile stimulation, the looming sounds were located at different distances; this was accomplished by delivering the tactile stimulus at different temporal delays starting from sound onset (T_{before}, T1, T2, T3, T4, T5, T_{after}). The sound source location at each temporal delay condition is indicated by triangles (black triangles for the left hemi-space and white triangles for the right hemi-space).

Emotional evaluation task

In order to assess any habituation phenomenon and ascertain that participants actually perceived dog growling as threatening and sheep bleating as non-threatening, participants performed a short emotional evaluation task before and after the audiotactile test. The two auditory stimuli (non-spatialized) were presented through Sennheiser HD650 headphones; each stimulus was presented only once. The order of stimuli presentation was counter-balanced between subjects. Participants had their eyes closed during the display of the sounds. After the offset of the sound, participants had to indicate the perceived valence and arousal of the sound on a 10 cm visual analogic scale (VAS).

RESULTS

EMOTIONAL EVALUATION TASK

Participants' responses on the VAS were not normally distributed for each sound stimulus. Hence, we compared the valence and arousal scores between the two sound stimuli and between groups using non-parametric tests.

As **Figure 2** shows, both groups perceived the dog sound as more negatively valenced than the sheep sound in each emotional evaluation (Wilcoxon test: $T < 9.00$, $p < 0.003$ in all cases). The perceived valence of the dog sound was not different between groups before the audiotactile test (Mann–Whitney test: $U = 69.00$, $p = 0.074$) and was significantly more negative in the DF group than in the NF group after the audiotactile test ($U = 38.50$, $p = 0.002$). The perceived valence of the sheep sound

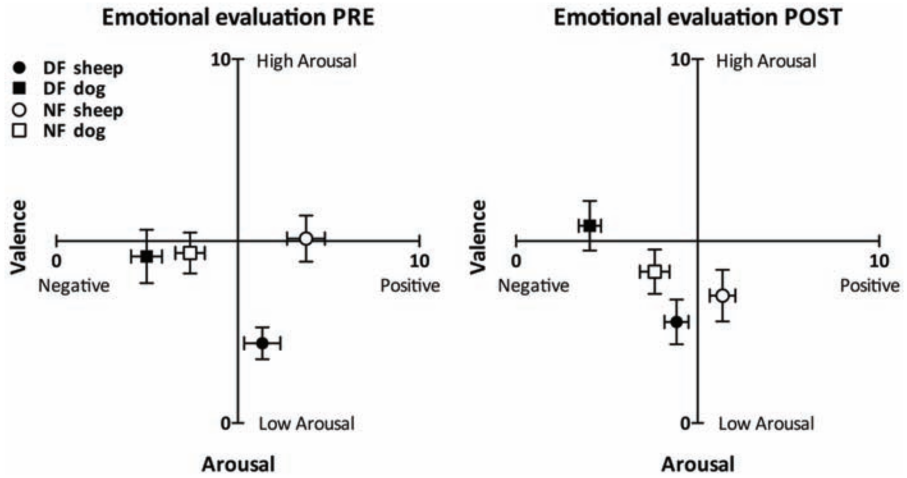

FIGURE 2 | Emotional evaluation task results. This figure depicts the perceived arousal and valence scores (mean ± SEM) reported by the dog-fearful (in black, $n_{DF} = 15$) and non-fearful (in white, $n_{NF} = 15$) groups in response to the non-threatening (circles) and threatening (squares) sounds, in the pre- (left) and post-audiotactile task (right) emotional evaluations. The perceived valence of the dog sound was more negative than the perceived valence of the sheep sound within each group and in both emotional evaluations. Moreover, within each group, while the sheep sound was rated as positive or neutral, the dog sound was rated as negative.

tended to be more positive in the NF group than in the DF group in both emotional evaluations ($U = 67.00$, $p > 0.058$ in both cases).

The DF group perceived the dog sound as more arousing than the sheep sound ($T < 20.00$, $p < 0.024$ in both emotional evaluations), while the NF group perceived the two sounds as similarly arousing ($T > 38.00$, $p < 0.211$ in both emotional evaluations). There was no significant difference of dog sound arousal scores between the NF and the DF group ($U > 77.00$, $p > 0.146$ in both emotional evaluations). As for the sheep sound, it was perceived as more arousing by the NF group compared to the DF group before the audiotactile test ($U = 35.00$, $p = 0.002$). After the audiotactile test, there was no more significant difference of sheep sound arousal scores between the NF and the DF group ($U = 77.00$, $p = 0.146$).

The results of this control test confirmed that the dog and the sheep sounds were respectively perceived as threatening and non-threatening in both the NF and the DF groups.

MAIN EXPERIMENT

Two participants (one NF and one DF) were excluded from the analyses because they perceived all the stimuli as coming from the frontal hemi-field. Two participants (DF) were also excluded because their mean RTs were substantially elevated, giving us reason to suspect that they did not correctly perform the task. As the rates of false alarms and omissions were very low – 0.38 and 0.58%, respectively – participants were extremely accurate in performing the task. Consequently, the performances were only analyzed in terms of RT. One participant (DF), however, had a high rate of misses (8.48%) and was therefore excluded from the RT analyses. The analyses on the audiotactile test were conducted on the 25 remaining participants ($n_{NF} = 14$; $n_{DF} = 11$). RTs non-precise measures due to interruptions from operating systems or device drivers were trimmed from the analyses. Mean RTs to tactile targets were calculated for each DELAY level and separately for each participant. RTs exceeding more than two standard deviations from the mean RT were considered outliers and also trimmed from the analyses (4.54% of the trials).

Mean RTs to tactile target were calculated for each of the 28 conditions (2 SOUND TYPE*2 HEMISPACE*7 DELAY). We first conducted an ANOVA on the mean RTs, with the between subject factor GROUP (NF/DF) and the within subject factors SOUND TYPE (threatening/non-threatening stimulus), HEMISPACE (left/right) and DELAY (T_{before}, T1, T2, T3, T4, T5, T_{after}). The global effect of DELAY was significant [$F_{(6,138)} = 31.42$, $p < 0.001$, $\eta_p^2 = 0.577$] suggesting that RTs were influenced by the time of tactile stimulation delivery. RTs in the unimodal condition T_{before} (391.69 ± 49.23 ms) were significantly slower than RTs in the bimodal conditions T1, T2, T3, T4, and T5 (*post hoc* Newman–Keuls' test: $p < 0.001$ in all cases). RTs in the unimodal condition T_{after} (353.92 ± 33.77 ms) were significantly faster than RTs at T_{before} (*post hoc* Newman–Keuls' test: $p < 0.001$). Given that RTs at T_{after} were significantly slower than RTs at T5 (*post hoc* Newman–Keuls' test: $p < 0.001$), we can exclude the possibility that participants were faster at late delays because of the increasing probability of receiving a tactile stimulation along trials. The difference in tactile RTs between T_{before} and T_{after} can be explained by the semantic content of the looming sounds, which places an animal in the environment; at T_{after}, participants potentially considered the animal as close to them but silent.

We then conducted an ANOVA on the mean RTs measured in the bimodal trials only, with the between subject factor GROUP (NF/DF) and the within subject factors SOUND TYPE (threatening/non-threatening stimulus), HEMISPACE (left/right) and DELAY (T1, T2, T3, T4, T5). The global effect of DELAY was significant [$F_{(4,92)} = 18.24, p < 0.001, \eta_p^2 = 0.442$]. The three-way interaction GROUP*SOUND TYPE*DELAY was also significant

[$F_{(4,92)} = 4.853$, $p = 0.001$, $\eta_p^2 = 0.174$] suggesting that RTs were differently modulated in the NF and the DF group depending on the perceived position of sound in space and as a function of whether the auditory stimulus was threatening or not. In the threatening condition, DF group's RTs were significantly faster when the tactile stimulus occurred at T2, T3, T4, and T5 compared to when the tactile stimulus occurred at T1 (*post hoc* Newman–Keuls' test: $p < 0.001$ in all cases). Contrastingly, in the non-threatening condition, DF group's RTs were faster when the tactile stimulus occurred at T4 and T5 compared to when it occurred at T1, T2, and T3 (*post hoc* Newman–Keuls' test: $p < 0.05$ in all cases). RTs at T2 were faster in the threatening condition compared to the non-threatening condition (*post hoc* Newman–Keuls' test: $p = 0.038$). RTs were not different between the threatening and non-threatening condition for the longest delays, i.e., closest distances (T3, T4, and T5) or for the smallest delay T1, i.e., the greater distance (*post hoc* Newman–Keuls' test: $p > 0.217$ in all cases). These results suggest that, in the DF group, the threatening sound began to affect tactile RTs at further distances compared to the non-threatening sound. In both the threatening and non-threatening condition, the NF group's RTs were significantly faster when the tactile stimulus occurred at T5 compared to when the tactile stimulus occurred at T1, T2, T3, and T4 (*post hoc* Newman–Keuls' test: $p < 0.002$ in all cases), suggesting that the distance at which the sound began to affect tactile RTs was similar in both the threatening and the non-threatening conditions.

In order to further investigate the influence of the different sounds on tactile RTs, we fitted participants' mean tactile RTs at the five delays with a sigmoid function using the same procedure as Canzoneri et al. The sigmoid function was described by the following equation: $y(x) = \frac{y_{\min} + y_{\max} \times e^{(x - x_i/b)}}{1 + e^{(x - x_i/b)}}$ where x represents the independent variable (i.e., the delay of tactile stimulation from sound onset in ms), y the dependent variable (i.e., tactile RT), $y_{\min}$ and $y_{\max}$ the lower and upper plateau of the sigmoid, x_i the value of the abscissa at the inflection point of the sigmoidal curve (i.e., the value of x at which $y = \frac{y_{\min} + y_{\max}}{2}$) and b is the slope at the inflection point. We estimated the parameters x_i and b for each participant's in each sound condition (threatening/non-threatening) and assigned *a priori* $y_{\min}$ and $y_{\max}$ to the minimum and maximum values of each data set. The sigmoid function better described participants' data than a linear function [$y(x) = y_0 \times x + a$, where y_0 is the intercept at $x = 0$ and a is the slope) as indicated by the result of the comparison of the root mean square errors ($RMSE_{sigmoid} = 7.80$ ms, $RMSE_{linear} = 8.69$ ms, Wilcoxon test: $T = 149.00$, $p = 0.001$). The parameter x_i was computed as a measure of the temporal delay, i.e., the distance, at which sound starts affecting tactile RTs and was analyzed in order to quantify PPS boundaries. As **Figure 3A** shows, DF group's x_i was lower in the threatening compared to the non-threatening condition [$t_{(8)} = -1.89$, $p = 0.030$, one-tailed, two participants were excluded due to bad fitting] suggesting that the boundaries of DF group's PPS in the threatening condition are farther from the participants than in the non-threatening condition. As **Figure 3B** shows, NF group's x_i did not significantly differ between sound conditions [$t_{(9)} = 0.19$, $p = 0.851$, two-tailed, four participants were excluded due to bad fitting] suggesting that the NF group's PPS size was similar in the threatening and in the non-threatening conditions. While the DF group's PPS was larger than the NF group's PPS in the threatening condition [$t_{(17)} = -2.73$, $p = 0.007$, one-tailed], there was no significant difference in PPS size between groups in the non-threatening

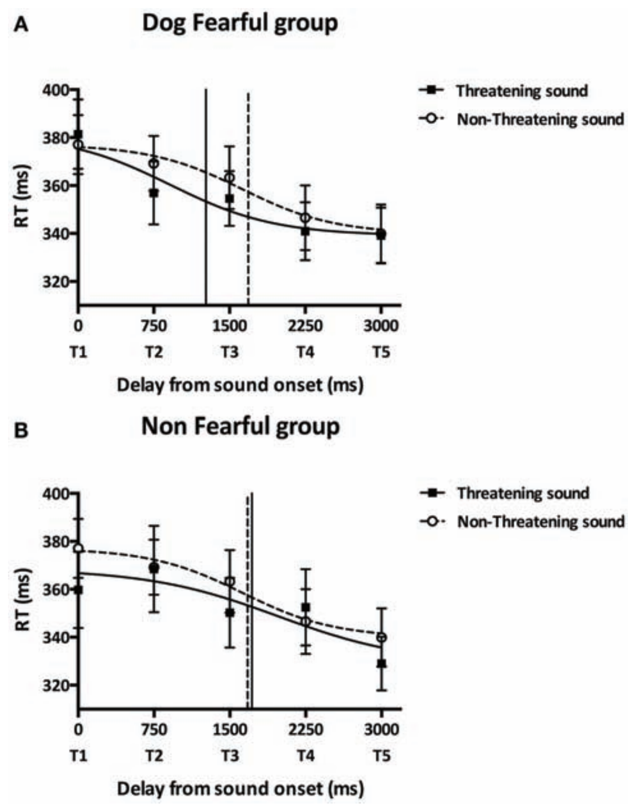

FIGURE 3 | Main experiment results. Participants performed the audiotactile task by responding to a tactile stimulation while a task-irrelevant threatening (dog growling) or non-threatening (sheep bleating) sound was looming toward them. This figure reports the mean tactile reactions times (±SEM) for the dog-fearful (top graph) and non-fearful group (bottom graph) in the threatening (black square) or non-threatening (white circles) sound conditions as a function of the delay of tactile stimulation delivery from sound onset. Reaction times were fitted with a sigmoid function. The inflection point abscissa of the sigmoid curves was computed as a measure of the temporal delay, i.e., the distance, at which sound starts affecting tactile RTs and was analyzed in order to quantify PPS boundaries. **(A)** Dog-fearful group results. The abscissa of the curve's inflection point was lower in the threatening sound condition (1266.81 ± 287.57 ms, black vertical line) compared to the non-threatening sound condition (1685.49 ± 548.41 ms, dashed vertical line) meaning that PPS boundaries were farther from participants in the presence of the dog sound than in the presence of the sheep sound. **(B)** Non-fearful group results. The abscissa of the curve's inflection point did not significantly differ between the threatening (1717.70 ± 413.23 ms, black vertical line) and the non-threatening (1675.15 ± 596.56 ms, dashed vertical line) sound conditions suggesting that participants' PPS size was similar in the presence of the dog and the sheep sounds. While the dog-fearful group's PPS was larger than the non-fearful group's PPS in the presence of the dog sound, there was no significant difference in PPS size between groups in the presence of the sheep sound.

condition [$t_{(17)} = 0.04$, $p = 0.485$, one-tailed]. Participants' difference between x_i in the non-threatening condition and x_i in the threatening condition, i.e., the extension of PPS boundaries, was not significantly correlated with trait anxiety scores ($r = 0.318$, $p = 0.184$).

DISCUSSION

Approaching unpleasant sounds trigger a particularly intense emotional response suggesting an activation of defensive responses (16). Previous results demonstrated that at distances wherein individuals non-sensitive to cynophobia still feel comfortable, a virtual visual looming dog triggers high discomfort for individuals sensitive to cynophobia (17). This variance in distance, together with PPS's proposed role of implementing a safety margin around the body, leads us to hypothesize that fear-object looming toward the body will expand PPS boundaries.

Consistently, our results suggest that looming feared elements extend PPS; the space that individuals consider as belonging to themselves enlarges when they perceive a feared object. This result seems consistent with previous results demonstrating that individuals underestimate the time at which a visual looming stimulus will collide with them when the stimulus is threatening (snakes, spiders, angry faces) compared to when it is non-threatening (butterflies, rabbits, neutral faces) (18, 19). Vagnoni et al. also show that this underestimation of time-to-collision is bigger for individuals who are fearful of the threatening stimulus; the size of the underestimation is linked to individuals' level of snakes- and spider-related anxiety. If PPS is extended, the distance between the feared object and PPS boundaries is smaller. Consequently, the encounter with PPS occurs sooner. Thus, the fact that an approaching feared stimulus is perceived as colliding sooner seems coherent with the PPS boundaries being farther.

Peri-personal space has also been shown as being extended after a satisfying social interaction (9). In our experiment, the expansion of PPS seems to aim at keeping unwanted and potentially harmful stimuli far from the body (i.e., outside PPS) and at allowing additional time for triggering defensive behaviors. In Teneggi et al. study, individuals' PPS boundaries did not enlarge in order to keep the other individual outside of PPS but rather to integrate them within it. In this case, the expansion of PPS would be linked to the implementation of approach behaviors.

Although PPS seems to be linked to emotional processes (20) and is thought to have a protective function, little is known on how PPS boundaries are influenced by anxiety. It has been shown that sensitivity to claustrophobic fear is related to larger PPS size as measured by a line bisection task (21). In their study, they observed a positive correlation between PPS size and the level of this space-related anxiety that is claustrophobic fear. This link was not observed with PPS size as measured by the hand-blink reflex defensive response (22). They instead observed a link between the size of PPS and trait anxiety. In contrast, results collected during a stop-approach task did not support a modulation of PPS size by anxiety (23). Our findings suggest that anxiety selectively influences PPS: sensitivity to cynophobia expands PPS boundaries when there is a dog stimulus in the environment. The diversity of results is potentially explained by the variety of experimental settings, which deliver different amount of fear-relevant features.

Though we studied a non-clinical sample, this situation-dependent effect of dog fear suggests that, at least in cynophobia, selective distortion of PPS is involved. Intrusion in PPS triggers high discomfort and regulative behaviors such as flight (5). When not constrained by the physical environment, individuals typically prevent undesired components of the environment from entering their PPS by adjusting their distance from them. Over-projecting PPS could allow more time to prepare defensive or avoidant behaviors in case of attack. The expansion of PPS in the presence of feared elements fits with the proposed protective function of PPS, i.e., assuring a margin of safety around the body (2, 5). What is perceived to be a disproportionate reaction from cynophobic individuals in the presence of dogs may be partially attributed to a normal reaction to the intrusion of an undesirable stimulus in an enlarged PPS.

Clinical psychology has implicitly used the notion of the influence of anxiety on PPS with the widely used Behavioral Assessment Test (BAT). This test is used to assess the level of fear of the patient in relation to a phobic object that is coming closer to him/her. When comparing the distance between the individual and the feared object at the beginning of therapy to the distance at the end of the therapy, the BAT serves as a measure of success [e.g., Ref. (24)]. A positive treatment outcome, as revealed by the BAT, probably reflects a change in the boundaries of PPS. The acceptable distance with the feared object is therefore a critical criterion in the assessment of severity of phobias. Our results suggest that PPS distortion could play a role in several phobias and that shrinking the oversized PPS could be a treatment strategy when facing fear-relevant situations.

Because anxiety regulation is shaped by the social context, we think it is important to take social distances into account when appraising anxiety mechanisms. Space is not a unitary construct in the brain and its neural representation is parceled across different compartments according to the behavioral interactions we have with them (25). Interactions between self and others can spread across the different compartments of space. It has already been suggested that space perception and representation might be distorted by anxiety [see Ref. (26) for a review]. While it is mainly the influence of anxiety on extra-personal space perception that has been studied [e.g., Ref. (27, 28)], it seems that PPS is another compartment of space that is distorted by anxiety.

ACKNOWLEDGMENTS

This research was supported by the EU FP7-ICT-2011-7 project VERVE (http://www.verveconsortium.eu/), grant no. 288910. The research leading to these results has received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06. We are grateful to Emmanuel Fléty for his help with the apparatus for tactile stimulation. We thank Olivier Warusfel and Thibaut Carpentier for their help on the elaboration of spatialized auditory stimuli through binaural rendering. We thank Cassandra Visconti for proof-reading this manuscript for American English spelling.

REFERENCES

1. Rizzolatti G, Fadiga L, Fogassi L, Gallese V. The space around us. *Science* (1997) **277**(5323):190–1. doi:10.1126/science.277.5323.190

2. Graziano MSA, Cooke DF. Parieto-frontal interactions, personal space, and defensive behavior. *Neuropsychologia* (2006) **44**(6):845–59. doi:10.1016/j.neuropsychologia.2005.09.009
3. Hall ET. *The Hidden Dimension*. New York: Doubleday (1966).
4. Hayduk LA. Personal space: an evaluative and orienting overview. *Psychol Bull* (1978) **85**(1):117–34. doi:10.1037//0033-2909.85.1.117
5. Aiello JR. Human spatial behavior. In: Stokols D, Altman I, editors. *Handbook of Environmental Psychology*. New York, NY: John Wiley & sons (1987). p. 389–504.
6. Bassolino M, Serino A, Ubaldi S, Làdavas E. Everyday use of the computer mouse extends peripersonal space representation. *Neuropsychologia* (2010) **48**(3):803–11. doi:10.1016/j.neuropsychologia.2009.11.009
7. Maravita A, Iriki A. Tools for the body (schema). *Trends Cogn Sci* (2004) **8**(2):79–86. doi:10.1016/j.tics.2003.12.008
8. Serino A, Bassolino M, Farnè A, Làdavas E. Extended multisensory space in blind cane users. *Psychol Sci* (2007) **18**(7):642–8. doi:10.1111/j.1467-9280.2007.01952.x
9. Teneggi C, Canzoneri E, Di Pellegrino G, Serino A. Social modulation of peripersonal space boundaries. *Curr Biol* (2013) **23**(5):406–11. doi:10.1016/j.cub.2013.01.043
10. Lloyd DM, Coates A, Knopp J, Oram S, Rowbotham S. Don't stand so close to me: the effect of auditory input on interpersonal space. *Perception* (2009) **38**(4):617–20. doi:10.1068/p6317
11. Lourenco SF, Longo MR. The plasticity of near space: evidence for contraction. *Cognition* (2009) **112**(3):451–6. doi:10.1016/j.cognition.2009.05.011
12. Tajadura-Jiménez A, Pantelidou G, Rebacz P, Västfjäll D, Tsakiris M. I-space: the effects of emotional valence and source of music on interpersonal distance. *PLoS One* (2011) **6**(10):e26083. doi:10.1371/journal.pone.0026083
13. Viaud-Delmon I, Znaïdi F, Bonneel N, Doukhan D, Suied C, Warusfel O, et al. Auditory-visual virtual environments to treat dog phobia. *The Seventh International Conference on Disability, Virtual Reality and Associated Technologies with ArtAbilitation 2008*. Porto, Portugal (2008). p. 119–24.
14. Spielberger CD, Gorsuch RL, Lushene PR, Vagg PR, Jacobs AG. *Manual for the State-Trait Anxiety Inventory (Form Y)*. Palo Alto, CA: Consulting Psychologists Press (1983).
15. Canzoneri E, Magosso E, Serino A. Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PLoS One* (2012) **7**(9):e44306. doi:10.1371/journal.pone.0044306
16. Tajadura-Jiménez A, Väljamäe A, Asutay E, Västfjäll D. Embodied auditory perception: the emotional impact of approaching and receding sound sources. *Emotion* (2010) **10**(2):216–29. doi:10.1037/a0018422
17. Taffou M, Guerchouche R, Drettakis G, Viaud-Delmon I. Auditory–visual aversive stimuli modulate the conscious experience of fear. *Multisens Res* (2013) **26**:347–70. doi:10.1163/22134808-00002424
18. Brendel E, DeLucia PR, Hecht H, Stacy RL, Larsen JT. Threatening pictures induce shortened time-to-contact estimates. *Atten Percept Psychophys* (2012) **74**(5):979–87. doi:10.3758/s13414-012-0285-0
19. Vagnoni E, Lourenco SF, Longo MR. Threat modulates perception of looming visual stimuli. *Curr Biol* (2012) **22**(19):R826–7. doi:10.1016/j.cub.2012.07.053
20. Kennedy DP, Gläscher J, Tyszka JM, Adolphs R. Personal space regulation by the human amygdala. *Nat Neurosci* (2009) **12**(10):1226–7. doi:10.1038/nn.2381
21. Lourenco SF, Longo MR, Pathman T. Near space and its relation to claustrophobic fear. *Cognition* (2011) **119**(3):448–53. doi:10.1016/j.cognition.2011.02.009
22. Sambo CF, Iannetti GD. Better safe than sorry? The safety margin surrounding the body is increased by anxiety. *J Neurosci* (2013) **33**(35):14225–30. doi:10.1523/JNEUROSCI.0706-13.2013
23. Dosey MA, Meisels M. Personal space and self-protection. *J Pers Soc Psychol* (1969) **11**(2):93–7. doi:10.1037/h0027040
24. Lang PJ, Lazovik AD. Experimental desensitization of a phobia. *J Abnorm Soc Psychol* (1963) **66**(6):519–25. doi:10.1037/h0039828
25. Previc FH. The neuropsychology of 3-D space. *Psychol Bull* (1998) **124**(2):123–64. doi:10.1037/0033-2909.124.2.123
26. Viaud-Delmon I, Venault P, Chapouthier G. Behavioral models for anxiety and multisensory integration in animals and humans. *Prog Neuropsychopharmacol Biol Psychiatry* (2011) **35**(6):1391–9. doi:10.1016/j.pnpbp.2010.09.016
27. Jones RB, Humphris G, Lewis T. Do agoraphobics interpret the environment in large shops and supermarkets differently? *Br J Clin Psychol* (1996) **35**(Pt 4):635–7. doi:10.1111/j.2044-8260.1996.tb01220.x
28. Kallai J, Makany T, Csatho A, Karadi K, Horvath D, Kovacs-Labadi B, et al. Cognitive and affective aspects of thigmotaxis strategy in humans. *Behav Neurosci* (2007) **121**(1):21–30. doi:10.1037/0735-7044.121.1.21

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 02 July 2014; accepted: 21 August 2014; published online: 03 September 2014.
*Citation: Taffou M and Viaud-Delmon I (2014) Cynophobic fear adaptively extends peri-personal space. Front. Psychiatry **5**:122. doi: 10.3389/fpsyt.2014.00122*
This article was submitted to Systems Biology, a section of the journal Frontiers in Psychiatry.

8. AUDITORY-VISUAL STIMULATION AND SENSITIVITY TO CROWD PHOBIA

Exploring the influence of multisensory stimulation on emotional experience according to the distance to the affective event with virtual reality

8.1. Introduction

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 et al., 2006) and with natural multisensory aesthetic events such as musical performances (Vines et al., 2011, 2006). We also showed that the emotional experience induced by dogs in subjects specifically fearful of dogs was increased when they could both see and hear the dogs (Taffou, Guerchouche, Drettakis, & Viaud-Delmon, 2013). It is possible that this influence of multisensory affective events on emotional experience depends on their spatial distance from the subject. Spatial distance and fear are indeed inextricably linked because close events represent more of a threat than events located farther away (Mobbs et al., 2007). It is thus possible that the emotional experience induced by multisensory affective stimuli is influenced by their location at close or far distances from the perceiver.

This study investigated the effect of auditory-visual aversive stimuli on negative emotional experience as a function of their distance from the perceiver. We used virtual reality (VR) to manipulate the sensory presentation and the spatial location of auditory-visual virtual crowds. Crowds are fear-relevant stimuli for humans and can be genuinely fearsome for a subset of individual sensitive to crowdphobia. We recruited a non-clinical sample of participants sensitive to crowdphobia. We also recruited a non-clinical sample of participants non-sensitive to crowdphobia as a control group: crowds are not fearsome for this subset of the population. The sensitivity to crowdphobia was assessed psychometrically by a questionnaire and, the behavioral and subjective components of the crowdphobic fear were assessed using Behavioral Avoidance Tests (BATs). Participants explored a virtual scene, in which they encountered virtual crowd stimuli presented through the auditory channel, the visual channel or both channels. During this experimental navigation in VR, they were asked to report their discomfort using Subjective Units of Distress when the crowds were located far from them (8m) or close to them (2m). We compared the discomfort induced by unimodal (visual or auditory) and bimodal (auditory-visual) crowd stimuli at each distance (close or far).

8.2. Sensitivity to Crowd Phobia

In order to recruit two categories of non-clinical participants – a group sensitive to crowdphobia and a group non-sensitive to crowdphobia – for the experimental navigation in VR, a characterization of individuals' respective levels of crowdphobic fear needed to be first established. Fear of crowd is a symptom found in diverse disorders such as agoraphobia, social phobia and Parkinson disease, for example. To our knowledge, there was, however, no psychometric tool specifically designed to assess the level of crowdphobic fear at the time of our experience.

8.2.1. Development of the Crowd Phobia Questionnaire (CP-Q)

Items in the Crowd Phobia Questionnaire (CP-Q) were composed in order to consider different aspects of an encounter with a crowd: (1) the sensory modalities through which the crowd is sensed (auditory, visual and/or tactile stimulation), (2) the mobility of the crowd of people (static, dynamic) and (3) the type of movement (unidirectional, random). The questionnaire consists of 15 items describing common situations in which there is a crowd of individuals such as "standing in a crowded subway train or bus" or “making your way through the crowd in a nightclub in order to join a group of friends”.

Individuals have to choose the proposition between four alternatives (no discomfort, slight discomfort, moderate discomfort, extreme discomfort) that best describes the intensity of discomfort they would experience in each of the 15 situations. Each item is scored as follows: no discomfort = 0, slight discomfort = 1, moderate discomfort = 2 and extreme discomfort =

3. The minimal total score on the questionnaire is 0, with a maximum of 45. The French version of the CP-Q can be found in the Annex section.

8.2.2. Selection of participants for the experimental navigation in virtual reality

The CP-Q was designed in order to establish a broad-spectrum evaluation of the sensitivity to crowdphobia with the ultimate goal of identifying individuals with high and low sensitivities to this phobia. We defined an individual as highly sensitive to crowdphobia when his/her score was inferior to the scores of 80% of the population (< 20th centile). Symmetrically, we defined an individual as having a low sensitivity to crowdphobia when his/her score was superior to the scores of 80% of the population (> 80th centile).

A sample of 228 individuals (mainly students, 121 women, age: 24.55 ± 5.32) completed the CP-Q. The repartition of individuals according to their score on the CP-Q is reported in Figure 8.1. We used these results to select individuals for the experimental navigation in VR. Individuals invited to participate in the study had a score either inferior to 6.4 (20th centile) or superior to 22 (80th centile).

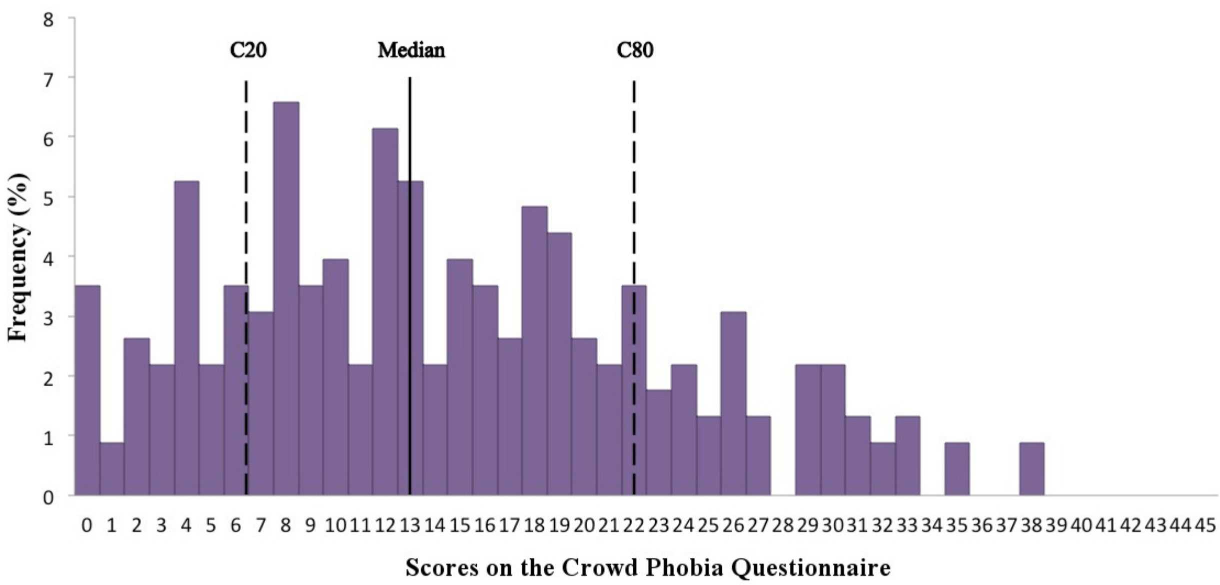

Figure 8.1. Frequency of scores on the Crowd Phobia Questionnaire.
The median score was 13 (black line) and the mean score was 14.61 (SD = 8.90). The 20th and 80th centile were respectively 6.4 and 22 (dashed lines).

8.3. Virtual environment containing crowds

8.3.1. Virtual reality setup

The VR setup was installed in an acoustically damped and soundproof recording studio. 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 (see Figure 8.2). Users wore polarized stereoscopic viewing 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 in different experiments involving binaural rendering (see Moeck et al., 2007; Sarlat, Warusfel, & Viaud-Delmon, 2006). With this procedure, the virtual sound source location can be manipulated by rendering accurate auditory cues such as inter-aural intensity and time differences and frequency spectrum. Ambient audio environment was 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 3D mouse to navigate in the virtual environment. With this device, they could control both rotations and translations within the virtual scene.

Figure 8.2. Virtual reality setup.
A user equipped with polarized glasses, headphones, a tracking device and a 3D mouse (left) and the stereoscopic passive screen (right).

8.3.2. Virtual environment containing crowds

8.3.2.1.Virtual environment

Virtual visual environment

The visual virtual environment (VE) we used was the Metropolis environment developed by the Graphics, Vision and Visualization (GV2) group of Trinity College Dublin, which was a partner in the VERVE project. Metropolis reproduces the outdoor environment of the Trinity College campus (See Figure 8.3) composed of buildings, alleys and vegetation. Animated virtual individuals, referred to as humanoids, can be placed in Metropolis. Different characteristics of these humanoids can be manipulated: gender, texture (eight different female and nine different male textures), animation (talking, listening), behavior (static, walking).

Figure 8.3. Metropolis visual virtual environment and humanoids.

Virtual auditory environment

The auditory virtual environment was developed within the Acoustic and Cognitive Spaces group at Ircam. The auditory virtual environment consisted of human speech and of an ambient audio environment composed of bird sounds and of urban activity.

We recorded the different sound files of human speech in the anechoic chamber at Ircam (See Figure 8.4). An anechoic chamber is a room designed to absorb as much sound as possible. The ceiling, floor and walls are covered with fiberglass wedge-shaped panels that absorb almost all acoustic energy. Sound therefore propagates in space without reflections. The speech sounds we obtained were recorded without their reflections, i.e. without the imprint of the room in which they were recorded. With this recording procedure, we can process resultant sound files with different rendering techniques in order to render sound in 3D and as though the sound source had been located in the environment of our choice (a small room, a church…).

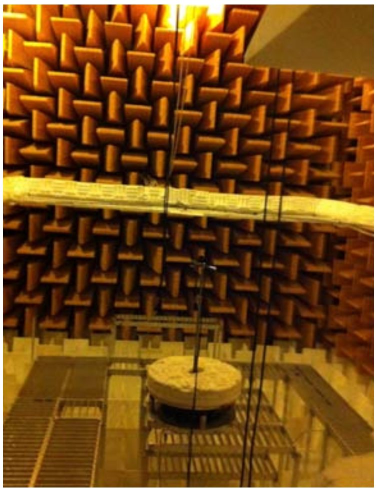

Figure 8.4. The anechoic chamber at Ircam

Twelve native French speakers (five women) participated in the recording session. In small groups of no more than four individuals, they were invited to enter the anechoic chamber and instructed to casually discuss with each other in French for five to ten minutes. The discussions were recorded with a MK6 Schoeps microphone. The resultant sound files were processed with Audacity software. Portions of sound files, which were either noisy or had several individuals talking at the same time, were removed. For each individual, sound files exclusively comprised of clipping of his/her speech were created. These sound files were then normalized and equalized in terms of loudness and compressed afterwards with standard voice compression parameters. Finally, the files were segmented in excerpts of different durations containing a sentence or an interjection.

8.3.2.2. Virtual crowd

The design of the virtual scene for the experimental navigation necessitated a definition of what constitutes a crowd in our VE: How many humanoids are needed for a group to be considered as a crowd? We created different groups comprising different amount of humanoids and conducted an experiment to determine which group to use for the experimental navigation in VR.

The participants were recruited independently from their scores on the Crowd Phobia Questionnaire (CP-Q) with the aim of selecting the groups of humanoids that would be considered as a "crowd" by the general population. Given that we planned to present auditory-only, visual-only and auditory-visual crowds in the experimental navigation in order to investigate the influence of sensory presentation on emotional experience, we had to select a group that is large enough to be considered as a crowd for each sensory condition.

During the experiment, participants completed a couple of questionnaires. Then, seven different groups of humanoids were presented to them during an immersion in virtual reality. We measured the extent to which they considered each group as a crowd. Participants evaluated the seven groups in three conditions of sensory presentation: auditory (A), visual (V) and auditory-visual (AV) presentation.

Methods

Participants

Twelve participants (2 women; age = 26.50 ± 4.60) with normal audition and vision voluntarily participated in the experiment. None of them had a history of psychiatric disorders, neurological disorders or was currently undergoing medical treatment. All participants provided written informed consent prior to the experiment, which was approved by the Health Research Ethics Committee (CERES) of Paris Descartes University.

Groups of humanoids

Seven groups were constructed (see Figure 8.5). The groups were composed of different amounts of humanoids (8, 16, 32, 48, 64, 96 and 128) organized in subgroups of 1 to 8. All of the humanoids were static and involved in a subgroup discussion as either a talker or as a listener. Talkers were attributed a talking animation and a gender-matching sound file of human speech whereas listeners were attributed only a listening animation. In order to avoid the technical difficulties of rendering interactive discussions between humanoids and to maintain the characteristics of stimulation stable over time, only one humanoid per subgroup was designated as a talker. Humanoids who were alone (subgroups of one individual), were talkers and talking in a mobile phone. The smallest group was composed of eight humanoids distributed among four subgroups. Then, the number of subgroups (and thus the number of talkers) was increased in parallel with the increase of the amount of humanoids with a ratio of one additional talker for each four additional humanoids. The groups of 16, 32, 48, 64, 96 and 128 humanoids were hence respectively composed of 6, 10, 14, 18, 26 and 34 subgroups.

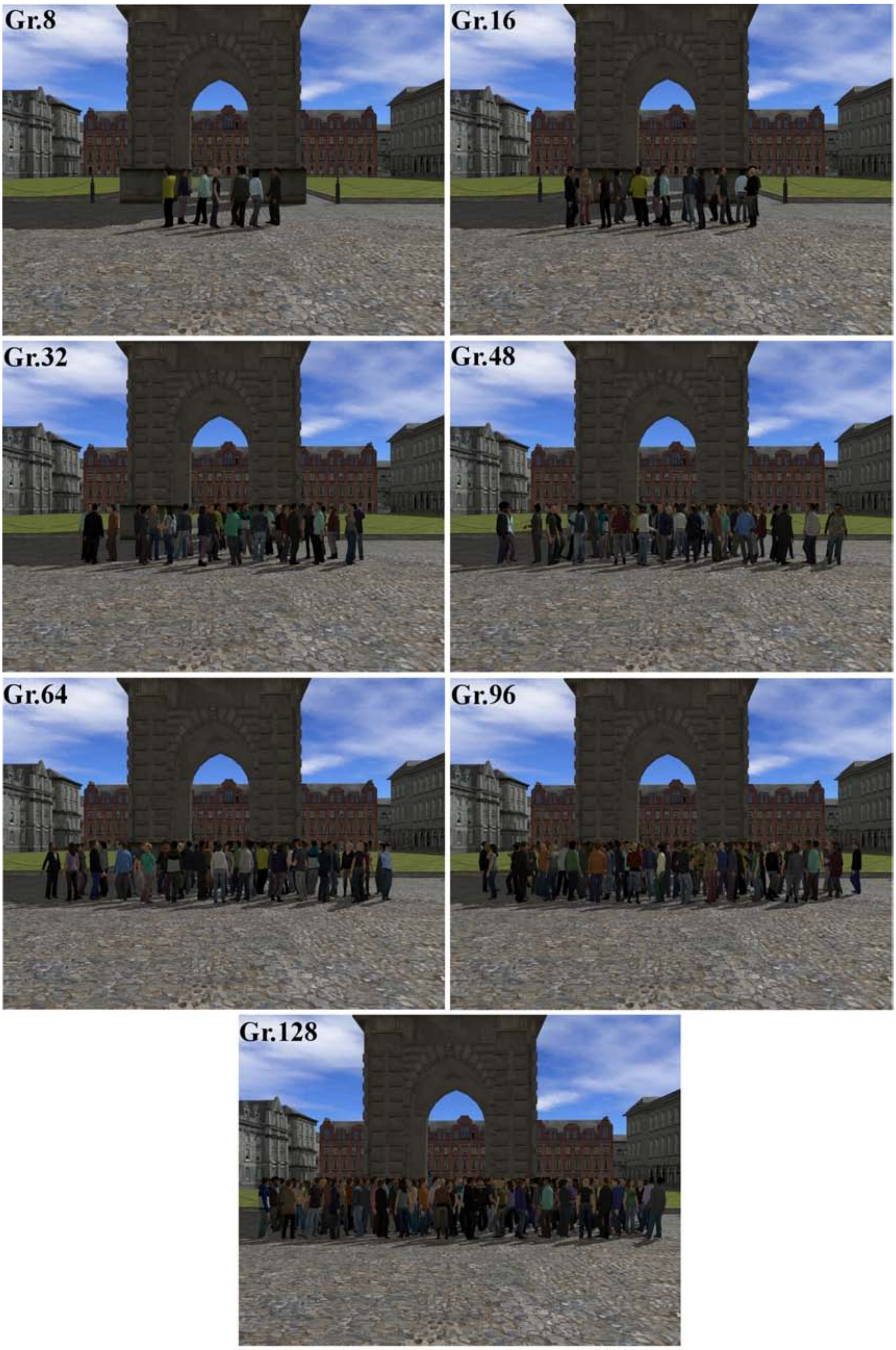

Figure 8.5. The seven different groups of humanoids.
The groups were composed of different amount of humanoids: 8 (Gr.8), 16 (Gr.16), 32 (Gr.32), 48 (Gr.48), 64 (Gr.64), 96 (Gr.96) or 128 (Gr.128) humanoids.

The position of the humanoids in space was defined so as that when standing at six meters from the group, a participant's field of view allowed all of them to be located. Groups were composed of an equal number of males and female with identical amounts of female talkers and male talkers. They were equally distributed in the right and left hemi-space of the user's field of view. To prevent the perception of a direct gaze, which could cause discomfort unrelated to the perception of the crowd per se, no humanoids were oriented toward the user.

Procedure

Upon arrival, participants completed the CP-Q and the trait portion of the State Trait Anxiety Inventory (STAI; Spielberger, Gorsuch, Lushene, Vagg, & Jacobs, 1983). Next, they had to complete three immersions in VR, during which the groups of humanoids were presented in three sensory conditions: auditory (A), visual (V) and auditory-visual (AV) conditions. During the immersions, participants were standing at 1.73m from the center of the screen (see section 8.3.1 for a description of the setup). The virtual scene placed them in a square of the VE (see section 8.3.2.1 for a description of the environment), in front of a big arch (see Figure 8.6). They did not navigate in the virtual scene. Each immersion was composed of seven steps. At each step, a group of humanoids was presented between the participant and the arch, with the closest humanoids being at 6m from the participant. As the experiment progressed, so did the number of humanoids composing the group. Participants were instructed to imagine that they were to walk through the arch to reach a building, using the shortest way possible. They could localize the spatial position of the arch visually and also via auditory cues (a bell ringing at the top of the arch).

For each group of humanoids, participants' had to indicate how much they agree with the following statement: "*There is a lot of people*". They used a scale from 0 (*I totally disagree*) to 10 (*I totally agree*) with 5 corresponding to: *I neither agree nor disagree*. This measure was defined as the crowd index of the group of humanoids. They also indicated the intensity of discomfort they experienced at each step using Subjective Units of Distress (SUD; Wolpe, 1973). Each participant evaluated first the groups of humanoids in the A condition, then in the V condition and finally in the AV condition. This order was chosen in order to prevent participants from mentally visualizing the group of humanoids in the A condition. In the A condition, participants' perception of the groups of humanoids was restricted to only auditory information by obscuring their vision with a mask; perception of only visual information in the V condition was achieved by blocking their hearing with earplugs and muting the sound coming from the virtual scene. After the immersions, a debriefing interview was conducted to assess and record participants' impressions.

Figure 8.6. A participant immersed in the virtual scene used to select the crowd for the experimental navigation.
He is standing at 6m from the group of 96 humanoids and is equipped with polarized glasses, headphones and head trackers

Results

Participants mean CP-Q score was 11.33 (SD = 6.62) with a median score of 13.50 (range [1.00; 19.00]). Trait anxiety scores ranged from 29 to 51 with a mean score of 38.25 (SD = 9.22) and a median of 39.00.

We calculated the mean and the confidence interval 95% of the mean (CI95) of the crowd indexes and the mean of the SUDs reported by participants for each group of humanoids (Gr.8, Gr.16, Gr.32, Gr.48, Gr.64, Gr.96, Gr.128) in each of the three sensory conditions (A, V and AV). We considered that a group of humanoids was a crowd when the lower boundary of the CI95 of the mean crowd index was higher than five. As shown in Figure 8.7, the groups of 96 and 128 humanoids met this criterion in the three sensory conditions. Participants' mean

SUDs in response to Gr.96 and Gr.128 in each of the three sensory conditions are reported in Table 8.1.

Table 8.1

Discomfort intensity reported in presence of the groups of humanoids considered as crowds

Mean SUD (± SD) in the three sensory conditions (A, V and AV) of crowds presentation

Sensory condition	A	V	AV
Gr.96	37.50 ± 27.34	36.67 ± 30.25	39.58 ± 30.49
Gr.128	41.67 ± 28.95	44.58 ± 32.30	47.08 ± 33.74

Discussion/Conclusion

We chose Gr.96 to be the stimulus for the experimental navigation in VR because according to our criteria, this group was considered to be a crowd. We also could have chosen Gr.128. However, in order to preserve the rendering performance of our systems, the group considered as a crowd and with the minimum of humanoids to display was selected to be the crowd stimulus for the experimental navigation. The different groups were presented in a growing order in terms of humanoids numerosity. We think that a randomized presentation order would have also revealed Gr.96 as a crowd. The contrast between Gr.96 and smaller groups would have certainly increased the reported mean crowd index compared to the contrast between Gr.64 and Gr.96 that our presentation order allowed to use during the evaluation of the group size.

It is worth noting that imagining walking through the crowd stimulus Gr.96 was associated with reports of negative feelings from the participants. The reported discomfort supports the idea that this crowd stimulus can be considered as unpleasant.

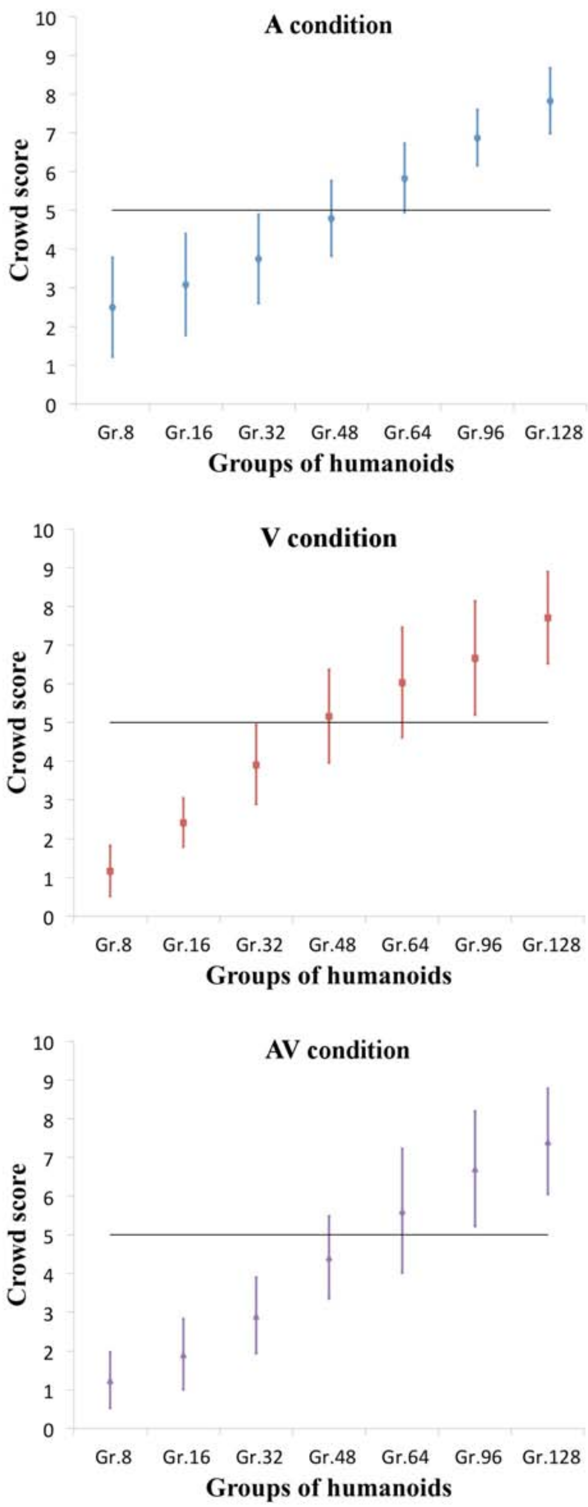

Figure 8.7. Results of the selection of the crowd for the experimental navigation.
Mean crowd index and CI95 for each group of humanoids in the auditory condition (A condition, top graph), visual condition (V condition, middle graph) and auditory-visual condition (AV condition, bottom graph). Two groups of humanoids met our criterion to be considered as a crowd (lower boundaries of CI95 higher than 5). Gr.96 and Gr 128 were considered as a crowd in each of the three sensory conditions of presentation.

8.4. Experimental navigation in virtual reality

The experimental navigation in VR aimed at testing the influence of the sensory modality (unimodal or bimodal) and the spatial location (close or far) of crowd stimuli on the negative emotional experience they induce.

8.4.1. Methods

First, participants were invited to take part in a twenty minute long diagnostic interview based on the Mini International Neuropsychiatric Interview. This interview was conducted to ascertain that no participant met criteria for pathological anxiety disorders and avoid any bias. Following this interview, participants were invited to complete several immersions in the VE. The total duration of the experimental session was two hours. The procedure was as follows: each participant was first submitted to three Behavioral Assessment Tests (BAT) in VR (see Mühlberger et al., 2008 for another example of a BAT conducted in VR) in order to assess the behavioral and subjective components of his/her crowdphobic fear according to the sensory modality, in which the crowd is presented (auditory, visual or auditory-visual). Then, before the experimental navigation in the auditory–visual VE, the participant became acquainted with the equipment and the navigation mode in a training immersion. The experimental navigation in the auditory–visual VE aimed at measuring negative emotional experiences in response to different sensory presentations of stimuli and as a function of the distance between the participants and the stimuli. Then, the participant was submitted a second time to the set of three BATs with the same procedure as the first time. Finally, he/she completed several questionnaires and was asked by the experimenter to comment on his experience (debriefing). All participants provided written informed consent prior to the experiment, which was approved by the Health Research Ethics Committee (CERES) of Paris Descartes University. Participants were paid 10€/hr.

8.4.1.1.Participants

Participants were selected on the basis of their scores on the Crowd Phobia Questionnaire (CP-Q see details in section 8.2.). Twenty-two healthy individuals (see details in Table 8.2) with normal audition and vision participated in the study. None of them had a history of psychiatric disorders, neurological disorders or was currently undergoing medical treatment. Ten individuals had a low score on the CP-Q and composed the NoFear group (NF). The remaining twelve individuals had high scores on the CP-Q and composed the CrowdFear group (CF). We used the State Trait Anxiety Inventory (STAI) (Spielberger et al., 1983) and the Liebowitz social anxiety scale (Liebowitz, 1987) to assess participants' anxiety levels. The Liebowitz social anxiety scale can be found in the Annex section.

Table 8.2

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
95% confidence interval	[21.81; 24.10]	[20.98; 24.02]	[21.47; 25.20]
CP-Q score ($M \pm SD$)[b]	16.64 ± 11.92	4.20 ± 1.81	27.00 ± 3.28
95% confidence interval	[11.35; 21.92]	[2.90 ; 5.50]	[24.92; 29.08]
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
95% confidence interval	[41.92; 49.54]	[34.90; 47.70]	[45.25; 53.58]
Liebowitz social anxiety scale			
Anxiety sub-score ($M \pm SD$)[d]	25.45 ± 12.38	16.60 ± 6.29	32.83 ± 11.38
95% confidence interval	[19.97; 30.94]	[12.10; 21.10]	[25.61; 40.06]
Avoidance sub-score ($M \pm SD$)[e]	20.48 ± 11.32	13.25 ± 5.36	26.50 ± 11.57
95% confidence interval	[15.46; 25.49]	[9.42; 17.08]	[19.15; 33.85]

[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 phobia scores significantly deviated from a normal distribution within each group, for which reason a non-parametric test was conducted. The crowd phobia score was significantly different between groups (Mann-Whitney U test: $U = 0.00$, $p < .001$).
[c]The trait anxiety scores was significantly different between groups (T test: $t_{(20)} = 2.45$, $p = .023$).
[d]The anxiety sub–score of the Liebowitz social anxiety scale was significantly different between groups (T test: $t_{(20)} = 4.02$, $p < .001$).
[e]The variance of the avoidance sub-scores was different between groups ($F(_9{}^{11}) = 4.66$, $p = .028$) 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: $U = 19.50$, $p = .014$).

Participants of the CF group presented higher scores of trait anxiety. They also had higher scores of social anxiety, which is consistent with the fact that the fear of crowds is a sub-component found in social phobia.

8.4.1.2.Virtual scenes and Virtual stimuli

During all immersions, participants were standing at 1.73m from the center of the screen (see section 8.3.1 for a description of the setup). The different virtual scenes presented during the participants' different immersions were designed within the Metropolis VE (see section 8.3.2 for a description of the environment).

The virtual scene used for all the BAT immersions was composed of a unique auditory-visual crowd stimulus located in a square of the VE, in front of a big arch. The crowd stimulus was the group of 96 humanoids (Gr.96, see Figure 8.5).

For the training immersion, the virtual scene was humanoid-free. Little numbered yellow flags served as beacons, tracing the path to be explored.

The virtual scene designed for the experimental navigation immersion was composed of several crowds, pairs of humanoids and solitary humanoids distributed in the VE. Three different crowd stimuli, based on Gr.96 stimulus (see Figure 8.5), were used: (1) an auditory stimulus, in which visual stimulation from the Gr.96 stimulus was blocked by a big flag, (2) a visual stimulus, in which auditory stimulation from the Gr.96 stimulus were prevented by depriving the group of the recordings of human speech and (3) an auditory-visual stimulus, which was the Gr.96 stimulus. Four copies of each stimulus (12 stimuli in total) were distributed in the VE so as to assure that, along the path to be explored, a stimulus of each sensory type (auditory, visual or auditory-visual) preceded, at least once, a stimulus of each other sensory type. Three stimuli consisting of a solitary humanoid, presented in the same conditions as the three crowd stimuli (A/V/AV), were also allocated along the exploration track. The order of presentation of all the stimuli is described in Table 8.3. Little numbered yellow flags were used as beacons to guide participants along the path to be explored. Little numbered red flags were positioned at 0.27m and 6.27m from the crowd stimuli. As participants stood at 1.73m from the screen, the distance between them and the crowd when they are at the red flags was respectively 2m and 8m. In order to facilitate the precision of the pause at the red flags, their texture changed, triggered by the participants' proximity to the target position. Three pairs of additional humanoids were added to fill the scene and increase realism.

Table 8.3

Order of stimuli presentation in the experimental 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

8.4.1.3.Questionnaires and interview measures

The state portion of the STAI (Spielberger et al., 1983) was used before and after completion of the total experimental protocol. 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 experimental navigation immersion. Discomfort ratings were collected during all of the immersions in the VE using the Subjective Unit of Distress (SUD;Wolpe, 1973).

8.4.1.4.Procedure

Figure 8.8 summarizes the procedure. Participants had completed the trait portion of the STAI several months before the experiment. The participants completed the Liebowitz anxiety scale and the state portion of the STAI upon arrival. Then they had to complete eight immersions in VR (set of three BATs, training, experimental navigation, set of three BATs).

Each participant was first invited to participate in the set of BATs pre-experimental navigation (BATs PRE). During these BATs, the participant was standing in a square at 10m from a crowd stimulus. The BAT was composed of ten steps. The first step was for the participant to face the crowd stimulus. Then, at each of the next eight steps, the participant was moved 1m closer in the VE to the crowd by the experimenter. For the final step, the participant had to approach the virtual crowd by taking a real step. At each of the ten steps, he/she had to rate his/her level of discomfort with SUDs. At each step, the experimenter proposed to stop the test should the participants was feel too anxious. If the participant agreed to continue, the next step started. The BAT score ranged from 0 to 10. Participants' score was 0 if they refused to face the crowd stimulus. The score 10 was attributed to participants who agreed to take a real step toward the crowd when standing at 2m from it. The set of BATs PRE consisted of three types of BATs: an auditory BAT (A BAT), a visual BAT (V BAT) and an auditory-visual BAT (AV BAT). Participants' perception of the crowd stimulus was restricted to only auditory information during the A BAT by obscuring their vision with a mask and to only visual information during the V BAT by blocking their hearing with earplugs and muting the sound coming from the virtual scenes. Given that it has been suggested that bimodal (auditory-visual) presentation of aversive stimuli evokes a more intense fear as compared to unimodal (auditory or visual) presentation of aversive stimuli (Taffou et al., 2013), participants went through A BAT and V BAT before AV BAT in order to avoid a saturation effect on fear. The A BAT was completed before V BAT so as to prevent participants to mentally visualize humanoids during A BAT. To recap, participants completed first the BAT A BAT, then the V BAT and finally the AV BAT.

In order to become acquainted with the equipment and the navigation mode, the participant went through a training immersion. During this immersion, participants were also trained to follow the path to be explored by using little 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 experimental navigation, which sought to measure participant's negative emotional experience at different distances from the auditory and visual crowd stimuli. During this immersion, participants had to explore the auditory-visual virtual scene. 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. If the participant was feeling too uncomfortable with being so close to the crowds, he/she was instructed to go as close as possible to the crowd and stop. Participants encountered the different crowd stimuli in the order described before (see section 8.4.1.2). 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, using SUDs, at 8m from the crowd (FAR condition) and then at 2m from the crowd (CLOSE condition). Four SUDs were collected in response to each of the crowd stimulus type (A, V, AV) and at each DISTANCE condition (CLOSE/FAR) for a total of 24 SUD measures.

After the experimental immersion, the participant completed the presence questionnaire from the I-group and the cybersickness scale. Then, he/she participated in a second set of BATs (BATs POST) with the same procedure used for the BATs PRE. He/she also completed a second state portion of the STAI. Finally, a debriefing interview was conducted to analyze and record the participant’s impressions.

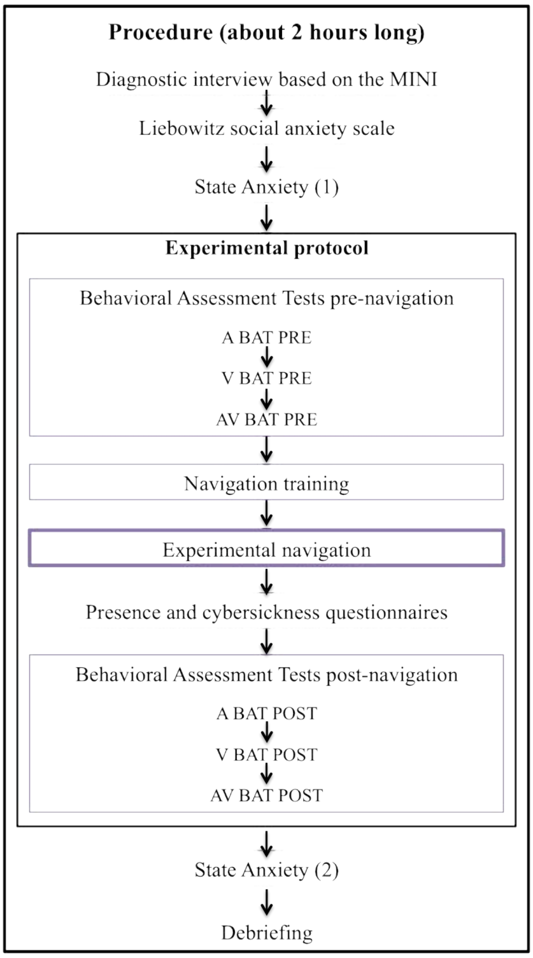

Figure 8.8. Procedure.
The immersions in virtual reality are framed in purple.

8.4.1.5.Data analyses

Questionnaire measures

We investigated possible differences in state anxiety scores between the NoFear and the CrowdFear groups and as a function of the moment at which they were recorded (at the beginning or at the end of the protocol) using an ANOVA. We conducted two-tailed non-parametric Mann–Whitney U tests to compare cybersickness and presence scores between groups.

Behavioral Assessment Tests (BATs)

We evaluated differences in the behaviorally-assessed crowdphobic fear between the CF and the NF group as well as possible differences in crowdphobic fear levels across BATs PRE and BATs POST by comparing several parameters resuming BAT results. The mean SUDs per step was calculated for each participant in each BAT immersion (A BAT PRE, V BAT PRE, AV BAT PRE, A BAT POST, V BAT POST, AV BAT POST). For each participant, we summed the SUDs they reported at each step of the BAT and divided the outcome by the number of step he/she managed to go through (i.e. BAT score).

In order to explore the spatial dynamic of experienced discomfort during the BATs, we studied a linear function to describe the relationship between the intensity of discomfort and the distance to the crowd. We used the distance to the crowd at each step of the BATs as the independent variable in our analyses. While the distance to the crowd at each of the first nine steps of the BATs was fixed in the virtual scenario (10m, 9m, 8m, 7m, 6m, 5m, 4m, 3m, 2m), the distance to the crowd at the final step of the BATs was assessed by measuring the step that participants made with a tape. In order to maintain the homogeneity of the independent variable during computation of the mathematical function, the SUDs collected at the final step of the BATs were excluded from the analyses. The linear function was described by the following equation: $y(x) = ax + y_0$; where x represents the independent variable (i.e. the distance to the crowd), y the dependent variable (i.e. SUDs), y_0 the value of y when $x = 0$, and a is the slope of the linear function. For each subject, the linear function was fitted, separately for each BAT, to the SUDs at the nine distances from the crowd in the least-square sense. The parameters a and y_0 were estimated during fitting and used to study the spatial dynamic of participants' discomfort during BATs. The parameter a was used as a reflection of the dynamic of discomfort increase as the distance to the crowd decreases. The lower (the more negative) a is, the faster discomfort increases along the approach of the crowd. The parameter y_0 was used as a reflection of the intensity of discomfort that would be experienced at very close distances from the crowd.

We separately analyzed the data for each sensory type of BAT (A/V/AV). For BAT PRE and POST, we compared the mean SUDs per step and the curve fitting parameters between groups (NF/CF) using two-tailed non-parametric Mann–Whitney U tests. We also compared, within each group, the mean SUDs per step and the curve fitting parameters between the sensory-paired BAT PRE and POST using two-tailed non-parametric Wilcoxon T tests for matched samples.

Experimental navigation

We tested the effect of bimodal crowd stimuli on the negative emotional experience as a function of their spatial location (close/far). 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). We tested the effect of GROUP on SUDs with two-tailed non-parametric Mann–Whitney U tests and the effect of SENSORY MODALITY (unimodal/bimodal) and DISTANCE (close/far) on SUDs using two-tailed non-parametric Wilcoxon T tests for matched samples.

We also tested the effect of the type of crowd stimulus (A/V/AV) on the negative emotional experience as a function of their spatial location (close/far). Within both groups (NF/CF), the mean SUDs reported in response to auditory (A-1, A-2, A-3, A-4), visual (V-1, V-2, V-3, V-4) and auditory-visual (AV-1, AV-2, AV-3, AV-4) crowd stimuli were calculated for both DISTANCE conditions (close/far). We tested the effect of GROUP on SUDs with two-tailed non-parametric Mann–Whitney U tests and the effect of CROWD STIMULUS (A/V/AV) and DISTANCE (close/far) on SUDs using two-tailed non-parametric Wilcoxon T tests for matched samples.

8.4.2. Results

One individual from the CF group (S01) 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).

8.4.2.1.Questionnaires measures

The state anxiety scores collected before and after the experimental protocol did not deviate from a normal distribution in the CF and the NF groups (Shapiro-Wilk test: $W > 0.90$, $p > .228$ in all cases). 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 8.4). The main effect of GROUP was significant ($F(1,19) = 8.25$, $p = .010$): state anxiety scores were higher in the CF group than in the NF group. There was no effect of neither the factor TIME ($F(1,19) = 0.112$, $p = .741$) nor the interaction GROUP*TIME ($F(1,19) = 0.148$, $p = .705$) on state anxiety scores.

Table 8.4

Individual Questionnaire Measures

ID Possible range	State anxiety 1 [20 –80]	State anxiety 2	Cybersickness [0 –88]	Presence [0 –84]
NoFear group				
NF-1	35	35	0	48
NF-2	29	27	1	38
NF-3	24	23	0	48
NF-4	27	24	2	47
NF-5	26	25	0	41
NF-6	29	31	1	36
NF-7	24	20	0	42
NF-8	23	23	1	32
NF-9	34	40	13	14
NF-10	27	31	10	38
M ± SD	27.80 ± 4.08	27.90 ± 6.24	2.80 ± 4.69	38.40 ± 10.11
CrowdFear group				
CF-1	45	47	4	43
CF-2	32	37	4	52
CF-3	26	28	7	36
CF-4	27	35	5	57
CF-5	51	21	20	48
CF-6	34	50	2	53
CF-7	38	32	3	53
CF-8	38	44	5	41
CF-9	28	28	17	41
CF-10	31	27	15	37
CF-11	42	27	1	55
M ± SD	35.64 ± 7.99	34.18 ± 9.37	7.55 ± 6.58	44.82 ± 11.10

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).

The presence scores revealed that participants globally had the feeling of being in the campus during the experimental navigation immersion. There was no difference in presence scores (U = 33.50, p = .139) between the two groups.

8.4.2.2.Behavioral Assessment Tests (BATs)

Among the 21 participants who completed the protocol, 18 reached the final step in each of the six BAT immersions and thus obtained maximal scores. The other three did not manage to get to the end of at least one BAT because of high discomfort. They all belonged to the CF group. CF-1 stopped after the 4th to the 6th step in each BAT immersion. CF-9 managed to reach the final step in every BAT except for the V BAT POST during which he stopped at the 9th step. CF-11 managed to reach the final step in every BAT immersion except for the V BAT PRE. Most of the participants got a maximal score for all the BATs, for which reason we consider this measure as not sensitive enough to reveal potential changes in crowdphobic fear levels between groups and between the BATs PRE and the BATs POST. We used the mean SUD per step as well as the y_0 and a parameters, estimated with the linear model, to investigate the differences in crowdphobic fear between groups and between BATs PRE and BATs POST.

A BATs

During the A BATs, the mean SUD per step (see Panel A of Figure 8.9) was significantly higher in the CF group compared to the NF group in both A BAT PRE and A BAT POST (U = 13.00, p = .003 in both cases). In both A BATs, the estimated value of y_0 was higher in the CF group compared to the NF group (U = 14.00, p = .004 in both cases) and the estimated slope a was lower (more negative) in the CF group compared to the NF group (U < 26.00, p < .045 in both cases) suggesting that CF group's discomfort increased faster as the distance to the crowd diminished and reached higher level at very close distances from the crowd compared to NF group's participants. The mean linear curve (defined by the mean of participants' best fitting parameters) is plotted for each group and for both BAT PRE AO-I and BAT POST AO-I in Panel B of Figure 8.9.

Within each group, the mean SUDs per step and the estimated values of y_0 and a were not different between A BAT PRE and A BAT POST (p > .063 in all cases).

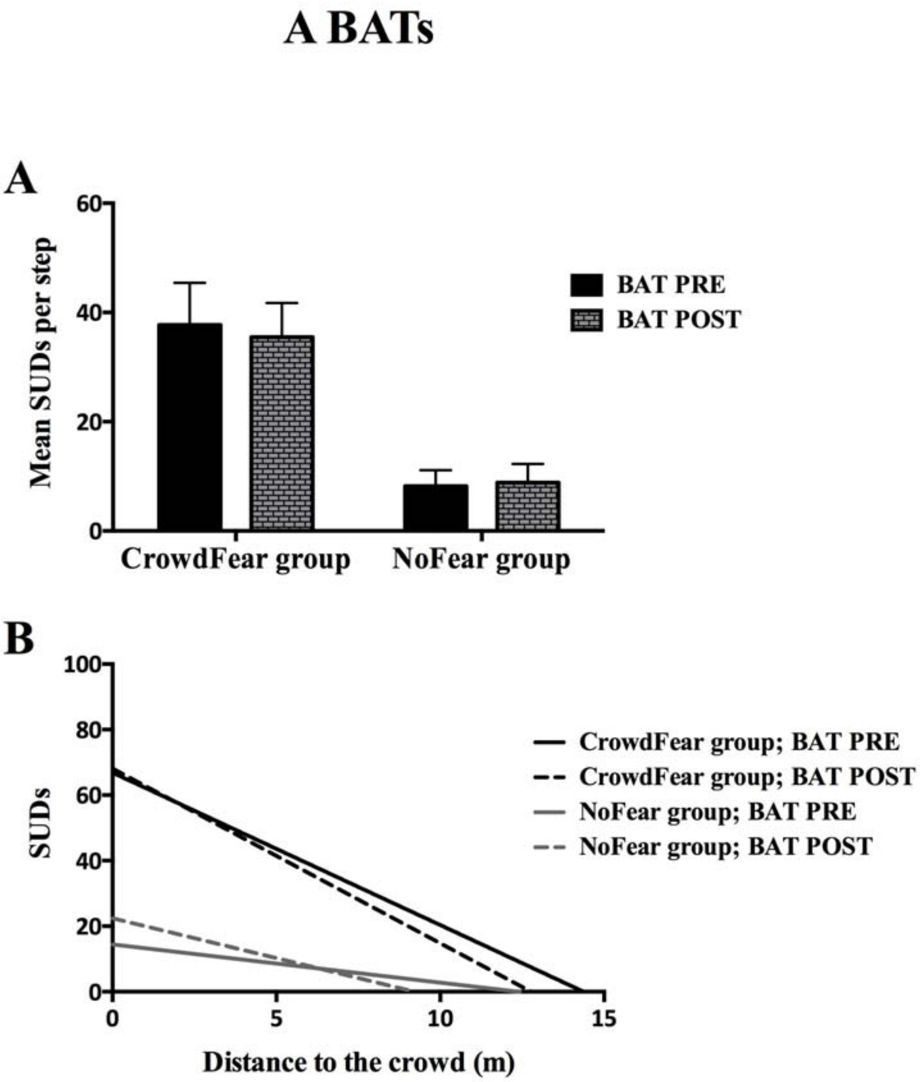

Figure 8.9. Auditory BATs results.
Panel A. Mean SUDs per step (± SE) reported during the auditory BAT pre- (black bars) or post- (checkered bars) experimental navigation by the CrowdFear group and the NoFear group. Panel B. Mean of the linear curves fitted to the data of the CrowdFear group (in black) and the NoFear group (in grey) during the auditory BAT pre- (regular lines) and post- (dashed lines) experimental navigation. The intersection point ordinate of the curves with the axis-ordinates was computed as a measure of the participant's discomfort upon contact with the crowd. The slope of the curves was used as a measure of the dynamic of discomfort increase along the approach towards the crowd.

V BATs

During the V BATs, the mean SUD per step (see Panel A of Figure 8.10) was significantly higher in the CF group as compared to the NF group in both V BAT PRE and V BAT POST ($U < 12.00$, $p < .003$ in both cases).

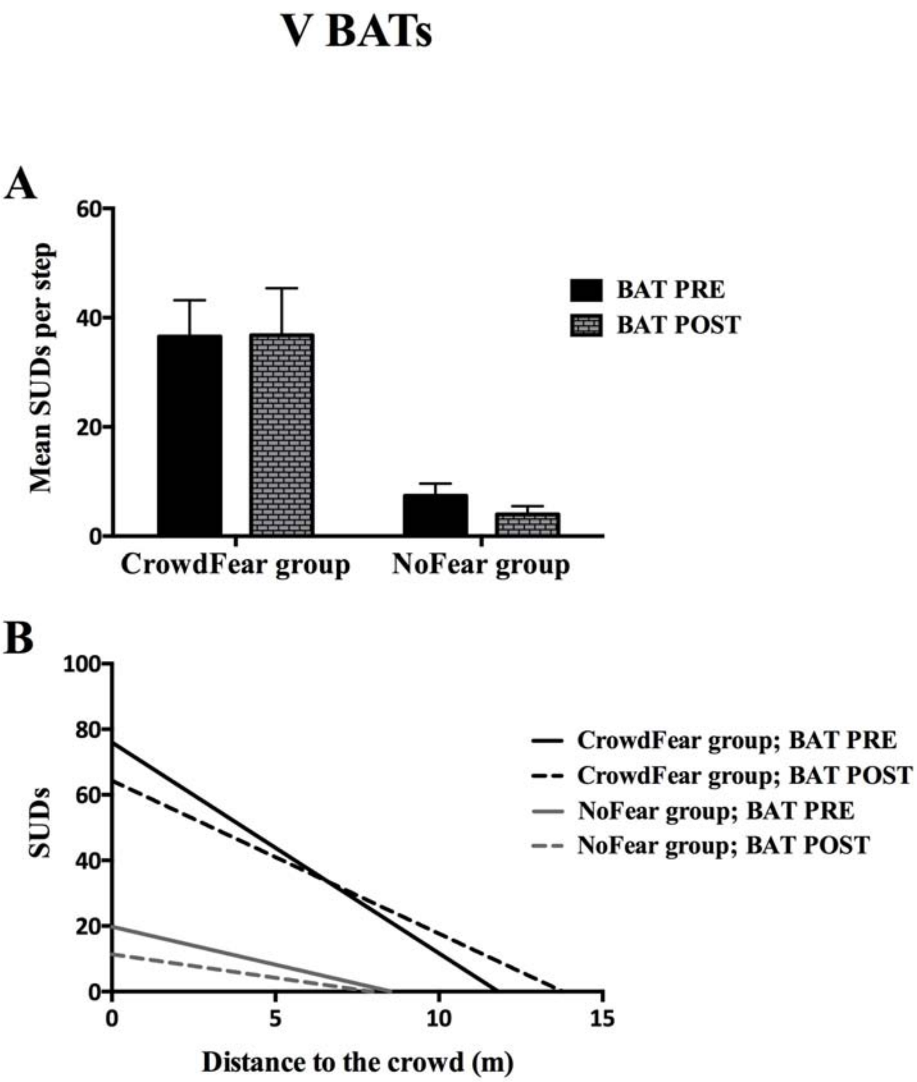

Figure 8.10. Visual BATs results.
Panel A. Mean SUDs per step (± SE) reported during the visual BAT pre- (black bars) or post- (checkered bars) experimental navigation by the CrowdFear group and the NoFear group. Panel B. Mean of the linear curves fitted to the data of the CrowdFear group (in black) and the NoFear group (in grey) during the V BAT pre- (regular lines) and post- (dashed lines) experimental navigation. The intersection point ordinate of the curves with the axis-ordinates was computed as a measure of the participant's discomfort upon contact with the crowd. The slope of the curves was used as a measure of the dynamic of discomfort increase along the approach towards the crowd.

In both V BATs, the estimated value of y_0 was higher in the CF group compared to the NF group (U < 13.00, p < .004 in both cases) and the estimated slope a was lower (more negative) in the CF group compared to the NF group (U < 19.00, p < .013 in both cases) suggesting that CF group's discomfort increased faster as the distance to the crowd diminished and reached higher level at very close distances from the crowd compared to NF group's participants. The mean linear curves for each group and for both V BAT PRE and V BAT POST are plotted in Panel B of Figure 8.10.

Within the CF group, the mean SUDs per step was not different between V BAT PRE and V BAT POST (T = 30.00, p > .790). The estimated y_0 in the V BAT POST was significantly lower than y_0 in the V BAT PRE (T = 9.00, p = .033) and the value of a was also significantly higher in the V BAT POST compared to the V BAT PRE (T = 4.00, p = .001). These results suggest that CF group's discomfort in the V BAT POST increased more slowly as the distance from the crowd decreased. Moreover, CF group's discomfort would reach lower level at very close distances from the crowd as compared to V BAT PRE. Within the NF group, the mean SUDs per step and the estimated values of y_0 and a were not different between V BAT PRE and V BAT POST (p > .116 in all cases).

AV BATs

During the AV BATs, the mean SUD per step (see Panel A of Figure 8.11) was significantly higher in the CF group compared to the NF group in both AV BAT PRE and AV BAT POST (U < 17.00, p < .009 in both cases). In both AV BATs, the estimated value of y_0 was higher in the CF group compared to the NF group (U < 17.00, p < .009 in both cases) and the estimated slope a was lower (more negative) in the CF group compared to the NF group (U < 26.00, p < .045 in both cases) suggesting that CF group's discomfort increased faster as the distance to the crowd diminished and reached higher level at very close distances from the crowd compared to NF group's participants. The mean linear curves for each group and for both AV BAT PRE and AV BAT POST are plotted in Panel B of Figure 8.11.

Within the CF group, the mean SUDs per step and the estimated y_0 were not significantly different between AV BAT PRE and AV BAT POST (p > .062). The estimated value of a was significantly higher in the AV BAT POST compared to the AV BAT PRE (T = 3.00, p = .008). These results suggest that CF group's discomfort in the AV BAT POST increased more slowly as the distance from the crowd decreased. However, the CF group's discomfort would reach a similar level at very close distances from the crowd than in the AV BAT PRE. Within the NF group, the mean SUDs per step and the estimated values of y_0 and a were not different between AV BAT PRE and AV BAT POST ($p > .398$ in all cases).

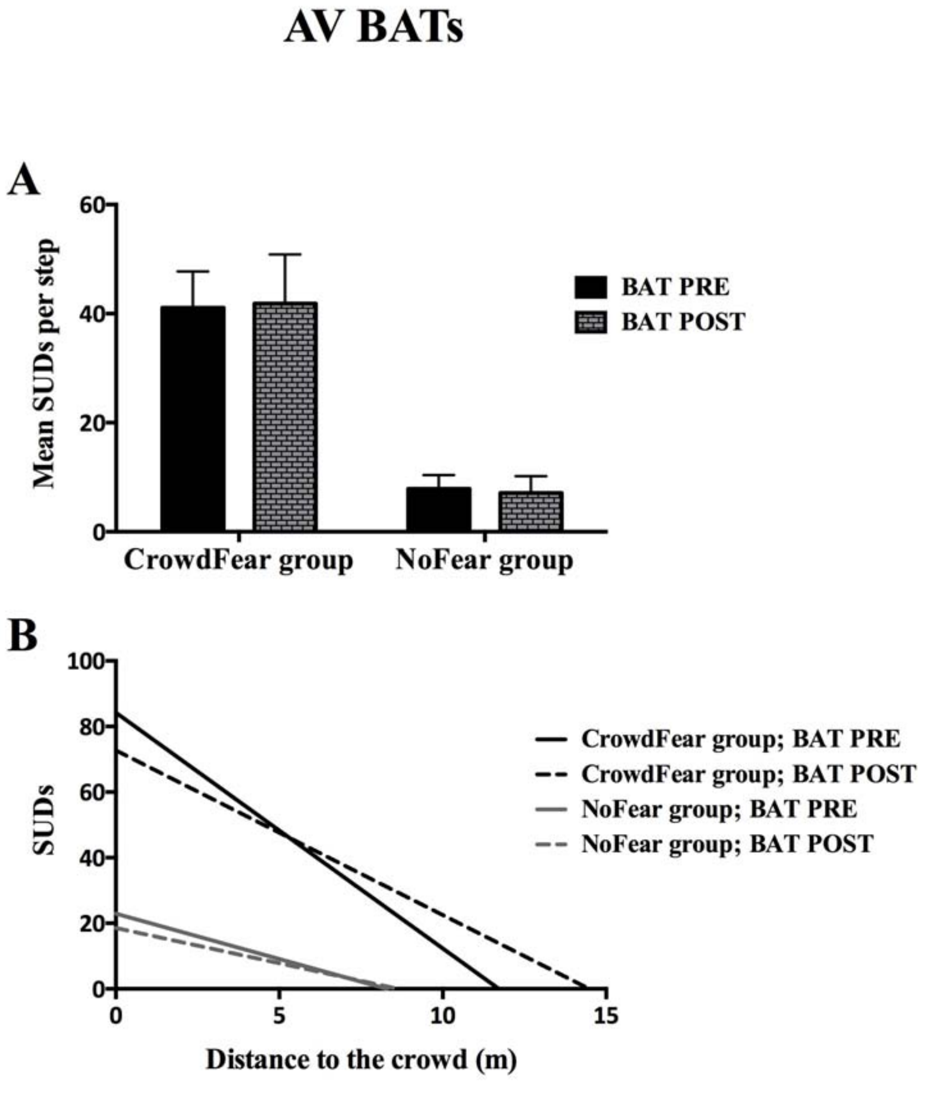

Figure 8.11. Auditory-visual BATs results.
Panel A. Mean SUDs per step (± SE) reported during the auditory-visual BAT pre- (black bars) or post- (checkered bars) experimental navigation by the CrowdFear group and the NoFear group. Panel B. Mean of the linear curves fitted to the data of the CrowdFear group (in black) and the NoFear group (in grey) during the auditory-visual BAT pre- (regular lines) and post- (dashed lines) experimental navigation. The intersection point ordinate of the curves with the axis-ordinates was computed as a measure of the participant's discomfort upon contact with the crowd. The slope of the curves was used as a measure of the dynamic of discomfort increase along the approach towards the crowd.

8.4.2.3. Experimental navigation

Each participant managed to come as close as 2m from each stimulus of the virtual scene.

Effect of the sensory modality of crowd stimulus on negative emotional experience

Effect of group

CF group's reported significantly higher SUDs than NF group's SUDs in response to the unimodal far (U = 13.00, p = .003), unimodal close (U = 14.00, p = .004), bimodal far (U = 11.50, p = .002) and bimodal close (U = 10.00, p = .002) crowd stimuli.

Effect of distance

SUDs in response to unimodal close crowd stimuli were higher than SUDs in response to unimodal far crowd stimuli in both the CF (T = 0.00, p = .003) and NF (T = 0.00, p = .018) groups. SUDs in response to bimodal close crowd stimuli were also higher than SUDs in response to bimodal far crowd stimuli in both the CF (T = 0.00, p = .003) and NF (T = 0.00, p = .043) groups.

Effect of sensory modality

In the distance condition CLOSE, as the left part of Figure 8.12 shows, CF group's SUDs were significantly higher in response to bimodal crowd stimuli compared to unimodal crowd stimuli (T = 0.00, p = .003). Contrastingly, NF group's reported SUDs were not different between the unimodal and bimodal conditions (T = 5.00, p = .128).

In the distance condition FAR, as the right part of Figure 8.12 shows, NF group's as well as CF group's reported SUDs were not different according to the sensory modality of the crowd stimuli (T > 3.00, p > .090 in both groups).

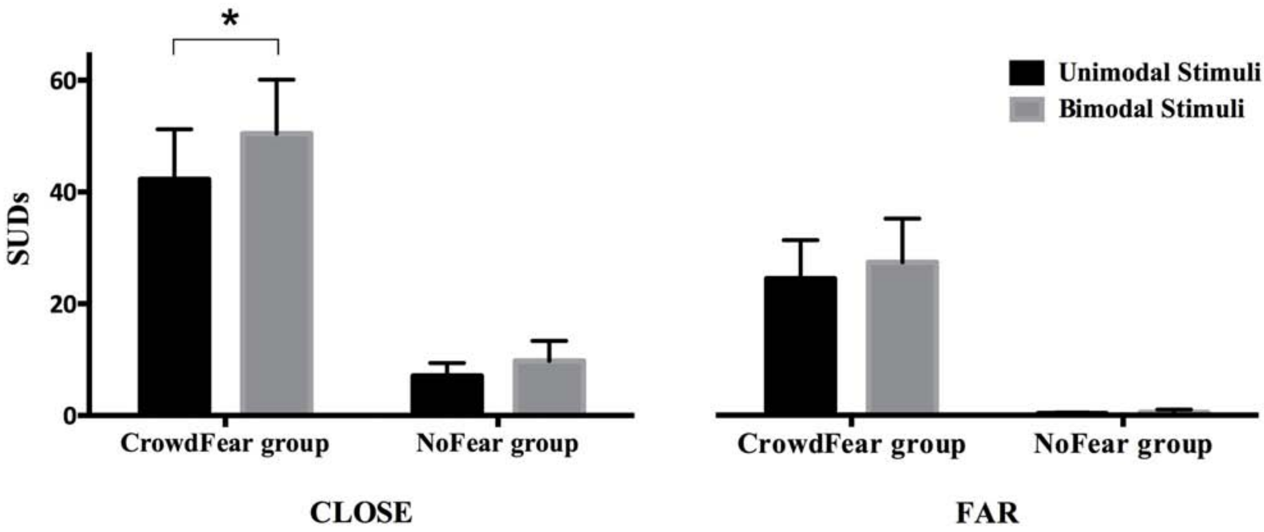

Figure 8.12. Effect of bimodal crowd stimuli on negative emotional experience.
Mean SUDs (± SE) reported by the CrowdFear group and the NoFear group in response to unimodal (auditory or visual; black bars) and bimodal (auditory-visual; grey bars) crowd stimuli at 2m (CLOSE) and 8m (FAR) distances during the experimental navigation. The bimodal crowd stimuli amplified the CrowdFear group's negative emotional experience when located at a close distance from participants.

Effect of the type of crowd stimulus on negative emotional experience

Effect of group

CF group's reported significantly higher SUDs than the NF group in response to each type of crowd stimulus (A/V/AV) in each the distance condition ($U < 16.50$, $p < .005$ in all cases).

Effect of distance

SUDs in response to auditory close crowd stimuli were higher than SUDs in response to auditory far crowd stimuli in both the CF ($T = 0.00$, $p = .005$) and NF ($T = 0.00$, $p = .028$) groups. SUDs in response to auditory-visual close crowd stimuli were also higher than SUDs in response to auditory-visual far crowd stimuli in both the CF ($T = 0.00$, $p = .003$) and NF ($T = 0.00$, $p = .043$) groups.

While SUDs in response to visual close crowd stimuli were higher than SUDs in response to visual far crowd stimuli in the CF group ($T = 0.00$, $p = .008$), there was no difference in the SUDs reported by the NF group in response to close and far visual crowd stimuli ($T = 0.00$, $p = .109$).

Effect of crowd stimulus

In the distance condition CLOSE, as the left part of Figure 8.13 shows, CF group's SUDs were significantly lower in response to visual crowd stimuli compared to auditory-visual crowd stimuli (T = 1.50, p = .008). There was no difference in SUDs reported by the CF group between auditory and auditory-visual crowd stimuli (T = 12.50, p = .126) or between auditory and visual crowd stimuli (T = 15.50, p = .120). NF group's reported SUDs tended to be lower in response to visual crowd stimuli compared to SUDs in response to auditory (T = 3.00, p = .063) and auditory-visual (T = 0.00, p = .068) crowd stimuli. There was no difference in SUDs reported by the NF group between auditory and auditory-visual crowd stimuli (T = 8.50, p = .353)

In the distance condition FAR, as the right part of Figure 8.13 shows, NF group's as well as CF group's reported SUDs were not different according to the type of the crowd stimulus ($T > 7.50$, $p > .142$ in all cases).

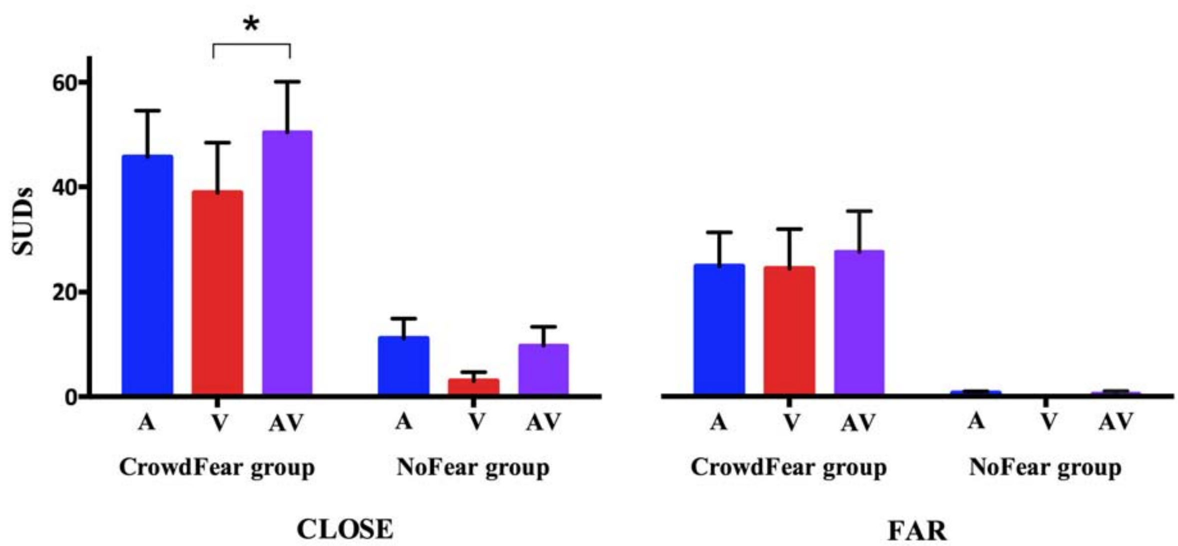

Figure 8.13. Effect of crowd stimulus' type on negative emotional experience.
Mean SUDs (± SE) reported by the CrowdFear group and the NoFear group in response to auditory (blue bars), visual (red bars) and auditory-visual (purple bars) crowd stimuli at 2m (CLOSE) and 8m (FAR) distances during the experimental 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. NF group's negative emotional experience tended to be lower in response to visual crowd stimuli than in response to auditory-visual crowd stimuli.

8.4.3. Discussion

The goal of this study was to examine how negative emotional experience is influenced by the sensory modality and spatial location of aversive stimuli. We used VR techniques to display crowd stimuli embedded in an auditory-visual VE and to control the characteristics of their presentation in terms of sensory and spatial parameters. We compared the level of discomfort induced by unimodal (auditory or visual) and bimodal (auditory-visual) crowd stimuli when they were located at close (2m) or far (8m) distances from healthy participants. We modulated the intensity of discomfort evoked by the crowd stimuli by recruiting two categories of participants: participants sensitive to crowdphobia [CrowdFear (CF) group] and participants non-sensitive to crowdphobia [NoFear (NF) group]. During the BATs, NF group's participants globally experienced less discomfort than CF group's participants. Moreover, whereas each participant of the NF group completed all the BATs, three participants of the CF group did not manage to complete at least one of the BATs. These behavioral and subjective results confirm the fact that the CF group considers crowd stimuli as more fearsome than the NF group and thus validate the use of these two groups to modulate the intensity of the negative experience induced by our crowd stimuli.

After the experimental navigation, a slight change was found in the AV BAT results of the CF group. The dynamic of the CF group's discomfort regarding the diminishing distances from the auditory-visual crowd was different after the experimental navigation containing crowds. CF group's discomfort increased less rapidly when approaching the auditory-visual crowd and reached lower intensity level at very close distances from the crowd. This suggests that after one immersion in our auditory-visual VE, the crowdphobic fear level of CF participants slightly diminished. This would not be surprising given that our procedure is inspired by protocols of virtual reality-based exposure therapy, which aim at treating anxiety disorders such as pathological fear (e.g. Botella et al., 1998; Emmelkamp, Bruynzeel, Drost, & Van der Mast, 2001; Garcia-Palacios, Hoffman, Carlin, Furness, & Botella, 2002; Riva, 2005; Rothbaum et al., 1995; Wald, 2004). This effect on AV BAT results seems to be linked to the changes observed in the BAT where only visual information from the crowd was available. The results of the BAT where only auditory information was available to the participants did indeed not differ before and after the experimental navigation. This might suggest that habituation to visual emotional cues is implemented faster than habituation in response to auditory emotional cues.

Distance and negative emotional experience

During the experimental navigation, CF participants' discomfort was higher in response to crowd stimuli located close to them than in response to crowds located far from them. A similar impact of distance on negative emotional experience induced by an aversive object has been put forward during social interaction (Schiffenbauer & Steven Schiavo, 1976). An unpleasant individual sitting at close distance induces a more negative experience compared to a seat farther away. Whereas NF participants experienced almost no discomfort at far distances from the crowd stimuli, they did report discomfort at 2m from the crowd. When located at close distances, the fear-relevant crowd stimuli may become fearsome for NF participants. Both CF and NF results are in-line with the fact that close events represent a greater threat than distant ones.

Sensory modality and negative emotional experience

Participants from the CF group 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 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; Taffou et al., 2013; Vines et al., 2011, 2006). Our findings further support the idea that the sensory modality of affective events influences conscious emotional experience.

When comparing the CF group's discomfort in response to auditory-visual crowds to the discomfort induced by each unimodal crowd type in the close distance condition, we observed that whereas visual crowds induced less discomfort than auditory-visual crowds, auditory crowds elicited similar intensity of discomfort than auditory-visual crowds. One possibility is that the emotional experience induced by bimodal auditory-visual crowds is controlled by whichever sensory cue that, on it's own, induces the most intense emotional experience. As such, if auditory cues elicit more intense feelings, as compared to visual cues, it follows that the intensity of the emotional experience in response to auditory-visual crowds would be equal to the one in response to auditory-only crowds. Under this hypothesis, there is no interaction between auditory and visual information in the production of the subjective affective experience; amplified feelings in response to bimodal auditory-visual crowds would then simply be linked to the multiplicity of sensory cues that are available.

Another possibility is that the increased discomfort in response to bimodal crowds as compared to unimodal crowds is linked to multisensory processes. Based on the literature regarding neural and behavioral responses to multisensory stimuli (see Alais et al., 2010 and Stein & Stanford, 2008 for reviews), responses to auditory-visual crowds would be expected to be higher than both the individual responses to visual and auditory crowds. However, we think that the subjective responses to affective stimuli that we collected here are less directly reflecting underlying multisensory integration processes than neural or speeded detection responses do. The production of the subjective emotional experience involves emotional regulation processes, which take context into account (Phillips et al., 2003). Even though participants received only auditory information when facing our auditory crowds, they are nevertheless aware that a visual component is linked to the auditory information but that, due to the big flag blocking their vision, they cannot perceive it. When facing our visual crowds, the participant's hearing of the crowd is not blocked and it is thus clear that no auditory component is linked to the visual crowd. This may explain the observed CF group's subjective reports of discomfort in response to the visual, auditory and auditory-visual crowds and does not necessarily discard the hypothesis that the increased discomfort in response to bimodal crowds compared to unimodal crowds is linked to multisensory processes. Visual and auditory inputs may interact to amplify discomfort. The amplified emotional experience in response to bimodal compared to unimodal crowds would then be linked to a subadditive cross-modal potentiation.

Sensory modality, distance and negative emotional experience

Of further interest is the fact that, when CF participants were standing at a farther distance from auditory-visual crowds, an increased discomfort was not found. At an 8m-distance, the intensity of participants' negative emotional experience was not influenced by the sensory modality (unimodal/bimodal) or the type (A/V/AV) of the crowds. Two explanations can be proposed to account for this differential influence of stimulus' sensory modality in accordance with the spatial distances from the perceiver. First, the difference may be linked to the intensity of the emotional experience induced by the auditory crowd stimulus at close and far distance conditions.

Second, the difference in the impact of the sensory modality on emotional experience at close and far distances could be linked to multisensory integration processes and explained by a specific multisensory processing of emotional cues located close to the body. As closeness increases, so does the quantity of sensory information received from external events. Moreover, the combination of different sensory emotional cues conveyed by an event contributes to improve the identification of its emotional significance in terms of accuracy and rapidity (e.g. Diederich & Colonius, 2004; Molholm, Ritter, Javitt, & Foxe, 2004; Ngo & Spence, 2010; Serino, Pizzoferrato, & Làdavas, 2008; Suied, Bonneel, & Viaud-Delmon, 2009). Thus, given that close events represent more of a threat than distant ones, specific multisensory integration processes dedicated to the emotional cues conveyed by affective events located close to the body would be particularly relevant. Many studies have brought evidence for a strong multisensory integration of neutral cues of 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 stronger auditory-tactile or visuo-tactile integration of neutral cues when both cues were located close to the body, as compared to farther distances. Our findings may be related to a similar stronger auditory-visual integration for emotional cues present at a close versus far distance.

The NF group's discomfort was not influenced by the sensory modality of the crowd in either the close or far condition. In the close condition, participants stood 2m from the crowd. Several studies have shown that fear level can modulate the perception of distances and remap the representation of far distances as close distances (Lourenco, Longo, & Pathman, 2011; Taffou & Viaud-Delmon, 2014; Vagnoni et al., 2012). It is thus possible that whereas 2m was already considered to be a close distance for the CF participants, it was still a far distance for the NF participants. It is thus logical that a strong multisensory affective process specific to events located at close distances would not be involved when NF participants stood at 2m from the crowds.

8.4.4. Conclusion

This study provides further evidence that the negative emotional experience induced by feared stimuli in individuals sensitive to phobia is enhanced by bimodal auditory-visual stimuli. This effect selectively occurred when the crowd was at a close, not at a far, distance from participants. These findings could help refine the design of VEs for the treatment of phobias. They indicate that combining the manipulation of sensory and spatial characteristic of feared situations can help address the disrupted affective processing in these anxiety disorders. Future work should investigate whether the specific increased emotional experience in response to closely-located multisensory affective stimuli is due to differences in the individual processing of sensory affective cues or to differences in multisensory processing of affective cues according to the distance of the affective events.

9. General Discussion

The present research work aimed at further understanding human affect. Three studies were conducted to investigate how negative feelings induced by feared objects are influenced by their sensory presentation, in virtual reality. This work showed that:

- there is a close relationship between affect and space
- bimodal aversive stimuli induce a more intense negative emotional experience
- there is an interaction of spatial and sensory influences on emotional experience

After a discussion of these three different aspects, different proposals as to how they can contribute to the field of virtual reality-based therapies will be presented.

Affect and space

The findings of the studies revealed a close relationship between affect and space. The second study demonstrated that unreasonable fear impacts the representation of spatial distances. In the presence of an especially feared object, far distances are remapped and represented as close distances. The third study showed that the location of aversive events at close or far distances influences the intensity of the negative emotional experience induced in the perceiver. Negative feelings were more intense in response to close rather than far aversive events. These findings are coherent with the survival-related function of affect. An aversive stimulus located at a close distance from the body represents a higher potential threat for life than a distant aversive stimulus. Increased negative feelings in response to close aversive events reflect a higher activation of the defensive system in charge of implementing appropriate defensive behaviors. The effect of unreasonable fear on the representation of close distances can also be related to the activation of the defensive system in the presence of a feared object. If an aversive stimulus represents a particularly high threat, it is safer to implement defensive behavior sooner. These results are coherent with previous work suggesting that a specific cerebral representation of the space close to the body serves a protective role, allowing for the preparation and implementation of defensive behaviors (Graziano & Cooke, 2006).

Affect and sensory modality

Both the studies with dogs and crowds revealed a more intense negative emotional experience in response to bimodal aversive stimuli as compared to unimodal stimuli. However, whereas the bimodal enhancement in the study with dogs was supraadditive, the bimodal enhancement in the study with crowds was subadditive. This difference in the size of the bimodal enhancement may be linked to the fact that the dogs were reactive to participants whereas the crowds were not. The dog stimuli's gaze were oriented towards participants and some of them started to growl and/or stood up when participants approached. In contrast, crowd stimuli did not react to the approach of participants; the humanoids were animated but immobile and their gaze was not oriented towards participants. Direct gaze has been shown to increase autonomic arousal in comparison to deviated gaze (e.g. Conty et al., 2010; Hietanen, Leppänen, Peltola, Linna-aho, & Ruuhiala, 2008). Moreover, motion has been shown to provide additional salience to threatening stimuli (Carretié et al., 2009). It is possible that these characteristics of the dog stimuli played a role in the size of the bimodal enhancement. Another possibility is that the difference in the size of bimodal enhancement with dogs and with crowds is linked to difference in the auditory stimuli distance to participants. In the study with dogs, the effect of the distance between the stimuli and the participant on emotional experience was not tested and the location of stimuli was not controlled. When the participants reported the intensity of their subjective feelings, all crowd stimuli were located at the same distance from them; this contrasts with the fact that the auditory dogs were located at greater distances from participants than visual and auditory-visual dogs at the moments of measure. Whereas auditory and visual dogs both induced weaker emotional experiences than auditory-visual dogs, auditory crowds induced similar intensity of emotional experience compared to auditory-visual crowds. The emotional experience induced by the auditory dogs would have certainly been more intense if they were located at the same distance as visual and auditory-visual dogs at the moment of measure. It is thus possible that we would have found similar results than with the crowds if we had controlled the distance between stimuli and participants in the study with dogs. Despite their differences, both studies suggest that beyond the early stage of affective processing, multisensory cues also modulate later stages of affective processing such as the production of an emotional experience.

Interaction of spatial and sensory influences on emotional experience

While bimodal aversive crowds increased negative feelings when located at distances close to the perceiver, the bimodal effect on emotional experience did not occur with distant aversive crowds. This effect might be linked to a specific multisensory affective process for stimuli located at close distances from the body and serve a survival-related purpose. Events represent indeed a greater threat when located close to the body and spatial proximity often increases the quantity of sensory information received from events; the spatial and sensory characteristics of harmful events are thus critical indicators of their threat potential for the body. The required increased activation of the defensive system in response to harmful events positioned to threaten life might be achieved through a strong multisensory integration of aversive cues located at close distances. This strong multisensory integration for emotional features of close events would potentiate affective processing in order to implement behaviors proportionate to the threat.

Altogether, the findings of the three studies suggest a close relationship between multisensory processes, space and affect. Multisensory affective stimuli enhance conscious emotional experience when located at close distances. Neuroimaging studies could allow for the determination of whether this enhancement of emotional experience is linked, or not, to a combination of the different sensory affective information. Our findings suggest that the location of the event is an important factor to consider for future studies on emotional experience induced by multisensory aversive events. The location of the affective event at close or far distances must be tightly controlled, especially as the representation of close distances can be modulated by participant's anxiety.

While this research has focused on negative emotional experience, it would also be interesting to investigate whether a similar influence of sensory modality and spatial location is also found with positive emotional experience induced by attractive events. Given that close distances enable interaction, it is possible that affective processing would be potentiated in order to increase the activation of the appetitive system, which implements approach behaviors. However, the fact that unpleasant looming sounds increase emotional experience whereas pleasant looming sounds do not (Tajadura-Jiménez, Väljamäe, Asutay, & Västfjäll, 2010) argue rather against this hypothesis. As for the effect of looming movement, the effect of multisensory presentation of close affective events on emotional experience could be specific to the defensive system.

Potential applications of the results to virtual reality-based therapies for phobias

This research work brought empirical evidences of the interest of manipulating sensory and spatial characteristics of feared objects in order to modulate the intensity of exposure in virtual reality-based therapy for phobias.

The gradation of exposure in virtual reality-based therapies often involves participants gradually approaching the object of their fear. Our participants who were sensitive to crowdphobia experienced more intense negative emotional experience when they were close versus far to the virtual crowd, be it auditory, visual or auditory-visual. The findings of this work provide empirical confirmation that it is an effective strategy to increase the intensity of subjects' feelings of fear. Furthermore, this work suggests that, for each subject, the location of the boundaries between distances represented as close and far in the presence of the fear object should be assessed. The location of these boundaries may vary according to the subject's level of fear. This information is thus important in order to modulate exposure intensity by manipulating spatial parameters of fear object presentation.

Participants sensitive to cynophobia as well as participants sensitive to crowdphobia reported increased negative feelings in response to auditory-visual compared to only visual presentation of the object of their fear. Moreover, after one immersion in our auditory-visual virtual environments, we already observed slight changes in participants' level of cynophobic and crowdphobic fear. These findings experimentally confirm that adding sensory affective information is an effective strategy to increase the intensity of exposure and suggest that multisensory virtual environments are particularly effective for the treatment of phobias. Additionally, the results obtained in the virtual environment containing crowds, suggest that multisensory stimulation are useless in regards to distant events; for virtual environments targeting crowdphobia at least, unimodal stimulation seems to be sufficient to address the unreasonable negative feelings in response to distant fear objects.

This research work also suggests a particular potential of auditory stimulation for the treatment of phobias in virtual reality. On a behavioral auditory-tactile task, an auditory-only dog stimulus was sufficient to observe differences between participants sensitive to cynophobia and participants non sensitive to cynophobia. Moreover, the negative feelings of participants sensitive to crowdphobia in response to visual crowds were amplified if the crowd also conveyed auditory information. They also experienced similarly amplified negative feelings in response to crowds conveying both visual and auditory information and in response to crowds hidden behind an obstacle and conveying only auditory information. Finally, the results on the BATs measuring crowdphobic fear suggest that habituation to auditory emotional inputs may be slower to implement in comparison to habituation to visual emotional inputs. It is possible that the visual component of feared objects is not needed in virtual environments to treat phobias. In our virtual environment containing crowds, there actually was an auditory-visual crowd behind the obstacle however no participants verified. The display of the 96 visual humanoids composing each crowd consumes a lot of energy in terms of computer performance and requires a lot of developmental work. If only a sound source located behind an obstacle allows for an intensity of exposure similar to one with auditory-visual stimulation, it could be of great interest for therapies. Furthermore, it would be interesting to test if a visual virtual environment is needed at all. Navigating in an only auditory virtual environment wherein subjects would be exposed to the auditory component of their fear object may possibly be sufficient to treat phobias.

Altogether, the present research brings new information on affect and suggests that sensory and spatial factors are important variables to take into account in the investigation of affect. It also exposes virtual reality as a relevant tool for the study of affect. Virtual reality might help us to better understand affective processing by providing more ecological stimulation and thus allowing for the investigation of factors seemingly involved in everyday human affective experience, such as spatial and sensory factors. Moreover, the findings can be directly exploited for research on virtual reality-based treatment for emotional disorders.

ANNEXES

DOG PHOBIA QUESTIONNAIRE (FRENCH VERSION)

Homme / Femme **Age :**

Date :

Merci de mettre une croix (X) devant la réponse choisie

A) Avez-vous plus peur des chiens que la plupart des gens ?

Oui Non

B) La présence d'un chien provoque-t-elle de la peur ou un sentiment d'anxiété chez vous ?

Oui Non

C) Il y a-t-il une race de chien qui vous fait particulièrement peur et si oui, laquelle ?

D) La taille d'un chien a-t-elle un effet sur le niveau de votre peur?

Veuillez coter votre niveau de peur face aux situations décrites dans le tableau ci-dessous :

0. Pas du tout peur
1. Un peu peur
2. Moyennement peur
3. Extrêmement peur

Situations	**0**	**1**	**2**	**3**
1) croiser un chien en laisse dans la rue				
2) s'approcher à un mètre d'un chien en laisse dans la rue				
3) croiser un chien sans laisse dans la rue				
4) s'approcher à un mètre d'un petit chien sans laisse				
5) s'approcher à un mètre d'un gros chien sans laisse				
6) s'approcher à un mètre d'un chien allongé ou qui dort				
7) s'approcher d'un chien qui remue la queue				
8) s'approcher d'un petit chien qui aboie				
9) s'approcher d'un gros chien qui aboie				
10) s'approcher d'un chien avec une muselière				
11) un chien qui vient spontanément à votre contact				
12) caresser le chien d'une connaissance, d'un ami(e) ou de la famille				
13) caresser un petit chien que vous ne connaissez pas dans la rue				
14) caresser un gros chien que vous ne connaissez pas dans la rue				

CROWD PHOBIA QUESTIONNAIRE (FRENCH VERSION)

Identifiant sujet : ………… **Genre :**…… **Age :**…… **Date:** ……

Veuillez indiquer le niveau d'inconfort que vous ressentez face aux situations décrites dans le tableau ci-dessous :

0. Aucun Inconfort
1. Un peu d'inconfort
2. Inconfort moyen
3. Extrême inconfort

Situations	0	1	2	3
Assister à un concert de musique classique dans une grande salle pleine.				
Attendre des amis à l'entrée d'un bar bondé. Depuis votre position, vous pouvez entendre le brouhaha venant de l'intérieur du bar.				
Marcher le long du quai de la gare après la descente du train un jour de grande affluence.				
Se trouver à proximité du lieu de passage du cortège de la gay-pride. Depuis votre position, vous entendez la musique et le tumulte de la foule.				
Marcher dans une station de métro à l'heure de pointe.				
Attendre des amis à l'entrée d'une piscine municipale. Depuis votre position, vous entendez le vacarme provenant des bassins.				
Se tenir debout dans un métro ou un bus bondé.				
Chercher sa place dans le train un soir de grand départ.				
Se frayer un chemin dans une discothèque pour rejoindre un groupe d'amis.				
Assister à un événement sportif (match de football, championnat, tournoi…) dans les gradins d'un stade ou d'une grande salle de sport.				
Se trouver au sein du cortège d'une manifestation.				
Marcher dans une galerie marchande le premier jour des soldes.				
Se trouver dans la fosse lors d'un grand concert dans un stade.				
Avancer au sein d'une file d'attente très dense (entrée de spectacle, de musée, parc d'attractions…)				
Se déplacer dans un bar bondé.				

CYBERSICKNESS QUESTIONNAIRE (FRENCH VERSION)

Date :
Heure :
ID sujet :
Environnement virtuel :
Immersion n° :

0 = signe absent
1= signe présent mais léger
2= signe modérément présent
3= signe sévèrement présent
4= signe très sévèrement présent

Palpitations cardiaques	0	1	2	3	4
Pression dans la poitrine	0	1	2	3	4
Faiblesse dans les bras ou les jambes	0	1	2	3	4
Tension musculaire, muscles endoloris	0	1	2	3	4
Fourmillements, picotements ou engourdissement dans certaines parties du corps	0	1	2	3	4
Faiblesse générale	0	1	2	3	4
Difficulté à respirer, respiration courte	0	1	2	3	4
Sensation de chaleur ou de froid	0	1	2	3	4
Sensation que les choses tournent	0	1	2	3	4
Points devants les yeux	0	1	2	3	4
Vision trouble ou distordue	0	1	2	3	4
Tremblements, frissonnements	0	1	2	3	4
Douleur dans le bas du dos	0	1	2	3	4
Transpiration excessive	0	1	2	3	4
Sensation de pression dans les oreilles	0	1	2	3	4
Vertige	0	1	2	3	4
Nausée	0	1	2	3	4
Estomac dérangé	0	1	2	3	4
Noeud dans l'estomac	0	1	2	3	4
Boule dans la gorge	0	1	2	3	4
Gorge sèche	0	1	2	3	4
Maux de tête	0	1	2	3	4
Sous-total par colonne					
TOTAL					

PRESENCE QUESTIONNAIRE FROM THE I-GROUP (FRENCH VERSION)

Voici plusieurs propositions qui peuvent s'appliquer à l'expérience que vous venez d'avoir. Indiquez, s'il vous plait, si chacune de ces propositions s'applique ou non à votre expérience. Vous pouvez utiliser n'importe quelle graduation. Il n'y a pas de bonne ou de mauvaise réponse, seule votre opinion est importante.
Vous remarquerez que certaines questions se ressemblent. Ceci est nécessaire pour des raisons statistiques. Rappelez-vous que vous devez répondre à ces questions en vous référant seulement à l'expérience que vous venez juste d'avoir.

1. A quel point étiez-vous conscient du monde réel environnant alors que vous étiez en train de naviguer dans le monde virtuel ? (par exemple : sons, température de la pièce, présence d'autres gens, etc.) ?

Extrêmement conscient	❍	❍	❍	❍	❍	❍	❍	Pas conscient du tout
	-3	-2	-1	0 Modérément conscient	+1	+2	+3	

2. Comment le monde virtuel vous a-t-il semblé?

Complètement réel	❍	❍	❍	❍	❍	❍	❍	Pas du tout réel
	-3	-2	-1	0	+1	+2	+3	

3. J'ai eu la sensation d'agir dans l'espace virtuel plutôt que d'agir sur un quelconque mécanisme à l'extérieur de celui-ci.

Pas du tout d'accord	❍	❍	❍	❍	❍	❍	❍	Complètement d'accord
	-3	-2	-1	0	+1	+2	+3	

4. A quel point votre expérience dans l'environnement virtuel vous a-t-elle semblée cohérente avec votre expérience dans le monde réel ?

Pas cohérente	❍	❍	❍	❍	❍	❍	❍	Très cohérente
	-3	-2	-1	0	+1	+2	+3	

5. A quel point le monde virtuel vous a-t-il semblé réel ?

A peu près aussi réel qu'un monde imaginé	❍	❍	❍	❍	❍	❍	❍	Indistinguable du monde réel
	-3	-2	-1	0	+1	+2	+3	

6. Je ne me suis pas senti présent dans l'espace virtuel.

Pas senti présent	❍	❍	❍	❍	❍	❍	❍	Senti présent
	-3	-2	-1	0	+1	+2	+3	

7. Je n'étais pas conscient de mon environnement réel.

Pas du tout d'accord	❍	❍	❍	❍	❍	❍	❍	Tout à fait d'accord
	-3	-2	-1	0	+1	+2	+3	

8. Dans le monde généré par l'ordinateur, j'ai eu le sentiment "d'y être ".

Pas du tout	❍	❍	❍	❍	❍	❍	❍	Beaucoup
	-3	-2	-1	0	+1	+2	+3	

9. D'une certaine façon, j'ai eu l'impression que le monde virtuel m'entourait.

Pas du tout d'accord	❍	❍	❍	❍	❍	❍	❍	Tout à fait d'accord
	-3	-2	-1	0	+1	+2	+3	

10. Je me suis senti présent dans l'espace virtuel.

Pas du tout d'accord	❍	❍	❍	❍	❍	❍	❍	Tout à fait d'accord
	-3	-2	-1	0	+1	+2	+3	

11. Je faisais toujours attention à l'environnement réel.

Pas du tout d'accord	❍	❍	❍	❍	❍	❍	❍	Tout à fait d'accord
	-3	-2	-1	0	+1	+2	+3	

12. Le monde virtuel semblait plus réaliste que le monde réel.

Pas du tout d'accord	❍	❍	❍	❍	❍	❍	❍	Tout à fait d'accord
	-3	-2	-1	0	+1	+2	+3	

13. J'avais l'impression que j'étais juste en train de percevoir des images.

Pas du tout d'accord	❍	❍	❍	❍	❍	❍	❍	Tout à fait d'accord
	-3	-2	-1	0	+1	+2	+3	

14. J'étais complètement captivé par le monde virtuel.

Pas du tout d'accord	❍	❍	❍	❍	❍	❍	❍	Tout à fait d'accord
	-3	-2	-1	0	+1	+2	+3	

Liebowitz Social Anxiety Scale (French version)

L'Echelle de Liebowitz

Source M. Bouvard et J. Cottraux Protocoles et échelles d'évaluations en psychiatrie et psychologie, Masson	ANXIETE 0 = Aucune 1 = Légère 2 = Moyenne 3 = Sévère	EVITEMENT 0= Jamais 1= Occasionnel (0-33%) 2= Fréquent (33-66%) 3= Habituel (67-100%)
1. Téléphoner en public (P)	—	
2. Participer au sein d'un petit groupe (P)	—	—
3. Manger dans un lieu public (P)	—	—
4. Boire en compagnie dans un lieu public (P)	—	—
5. Parler à des gens qui détiennent une autorité (S)	—	—
6. Jouer, donner une représentation ou une conférence (P)	—	—
7. Aller à une soirée (S)	—	—
8. Travailler en étant observé (P)	—	—
9. Ecrire en étant observé (P)	—	—
10. Contacter par téléphone quelqu'un que vous ne connaissez pas très bien (S)	—	—
11. Parler à des gens que vous ne connaissez pas très bien (S)	—	—
12. Rencontrer des inconnus (S)	—	—
13. Uriner dans les toilettes publiques (P)	—	—
14. Entrer dans une pièce alors que tout le monde est déjà assis (P)	—	—
15. Etre le centre d'attention (S)	—	—
16. Prendre la parole à une réunion (P)	—	—
17. Passer un examen (P)	—	—
18. Exprimer son désaccord ou sa désapprobation à des gens que vous ne connaissez pas très bien (S)	—	—

19. Regarder dans les yeux des gens que vous ne connaissez pas très bien (S)	—	—
20. Faire un compte-rendu à un groupe (P)	—	—
21. Essayer de "draguer" quelqu'un (S)	—	—
22. Rapporter des marchandises dans un magasin (S)	—	—
23. Donner une soirée (S)	—	—
24. Résister aux pressions d'un vendeur insistant (S)	—	—
TOTAL	A = ——	E = ——
Deux scores : S = Interaction sociale	A.S = ——	E.S = ——
P = Performance	P.S = ——	P.S = ——

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.

Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current biology : CB*, *14*(3), 257–62. doi:10.1016/j.cub.2004.01.029

Alais, D., Newell, F. N., & Mamassian, P. (2010). Multisensory processing in review: from physiology to behaviour. *Seeing and perceiving*, *23*(1), 3–38. doi:10.1163/187847510X488603

Anderson, P., Rothbaum, B. O., & Hodges, L. F. (2003). Virtual Reality Exposure in the Treatment of Social Anxiety. *Cognitive and Behavioral Practice*, *10*, 240–242.

Avillac, M., Ben Hamed, S., & Duhamel, J.-R. (2007). Multisensory integration in the ventral intraparietal area of the macaque monkey. *The Journal of Neuroscience*, *27*(8), 1922–32. doi:10.1523/JNEUROSCI.2646-06.2007

Banks, M. S. (2004). Neuroscience : What You See and Hear is What You Get. *Current biology*, *14*(2), 14–16. doi:10.1016/j.cub.2004

Barrett, L. F. (1998). Discrete Emotions or Dimensions ? The Role of Valence Focus and Arousal Focus. *Cognition and Emotion*, *12*(4).

Bassolino, M., Serino, A., Ubaldi, S., & Làdavas, E. (2010). Everyday use of the computer mouse extends peripersonal space representation. *Neuropsychologia*, *48*(3), 803–11. doi:10.1016/j.neuropsychologia.2009.11.009

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. doi:10.1016/j.brainres.2005.12.065

Beauchamp, M. S., Lee, K. E., Argall, B. D., & Martin, A. (2004). Integration of auditory and visual information about objects in superior temporal sulcus. *Neuron*, *41*(5), 809–23.

Beauchamp, M. S., Yasar, N. E., Frye, R. E., & Ro, T. (2008). Touch, sound and vision in human superior temporal sulcus. *NeuroImage*, *41*(3), 1011–1020.

Beck, J. G., Palyo, S. A., Winer, E. H., Schwagler, B. E., & Ang, E. J. (2007). Virtual Reality Exposure Therapy for PTSD Symptoms After a Road Accident: An Uncontrolled Case Series. *Behavior Therapy*, *38*(1), 39–48.

Bertelson, P., & Aschersleben, G. (1998). Automatic visual bias of perceived auditory location. *Psychonomic Bulletin & Review*, *5*(3), 482–489.

Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception & psychophysics*, *29*(6), 578–84.

Bertelson, P., Vroomen, J., De Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & psychophysics*, *62*(2), 321–332.

Bohil, C. J., Alicea, B., & Biocca, F. A. (2011). Virtual reality in neuroscience research and therapy. *Nature reviews. Neuroscience*, *12*(12), 752–62. doi:10.1038/nrn3122

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. doi:10.1007/s00221-004-2005-z

Bos, J. E. (2007). Why cybersickness. In S. So, R. Cheung, E. Chow, J. Ji, & A. Lam (Eds.), *Proceedings of the The First International Symposium on Visually Induced Motion Sickness, Fatigue, and Photosensitive Epileptic Seizures (VIMS2007)* (pp. 11–17). Hong Kong: The Hong Kong University of Science and Technology.

Botella, C., Baños, R. M., Perpiñá, C., Villa, H., Alcañiz, M., & Rey, A. (1998). Virtual reality treatment of claustrophobia: A case report. *Behaviour Research and Therapy*, *36*(2), 239–246.

Botella, C., Baños, R. M., Villa, H., Perpiñá, C., & García-Palacios, A. (2000). Virtual reality in the treatment of claustrophobic fear: A controlled, multiple-baseline design. *Behavior Therapy*, *31*(3), 583–595. doi:10.1016/S0005-7894(00)80032-5

Botella, C., Bretón-López, J., Quero, S., Baños, R., & García-Palacios, A. (2010). Treating Cockroach Phobia With Augmented Reality. *Behavior Therapy*, *41*(3), 401–413.

Botella, C., Villa, H., Baños, R. M., Quero, S., Alcañiz, M., & Riva, G. (2007). Virtual Reality Exposure in the Treatment of Panic Disorder and Agoraphobia : A Controlled Study †, *175*, 164–175. doi:10.1002/cpp

Botella, C., Villa, H., Banos, R., Perpina, C., & Garcia-palacios, A. (1999). The Treatment of claustrophobia with virtual reality: Changes in Other Phobic Behaviors Not specifically treated. *Cyberpsychology & Behavior*, *2*(2), 135–141.

Bradley, M. M., Codispoti, M., Cuthbert, B. N., & Lang, P. J. (2001). Emotion and motivation I: Defensive and appetitive reactions in picture processing. *Emotion*, *1*(3), 276–298. doi:10.1037//1528-3542.1.3.276

Bradley, M. M., & Lang, P. J. (2002). Measuring Emotion: Behavior, Feeling, and Physiology. In R. D. Lane & L. Nadel (Eds.), *Cognitive Neuroscience of Emotion* (pp. 242–276). Oxford University Press.

Bremmer, F., Schlack, A., Shah, N. J., Zafiris, O., Kubischik, M., Hoffmann, K. P., … Fink, G. R. (2001). Polymodal motion processing in posterior parietal and premotor cortex: A human fMRI study strongly implies equivalencies between humans and monkeys. *Neuron*, *29*(1), 287–296.

Brendel, E., DeLucia, P. R., Hecht, H., Stacy, R. L., & Larsen, J. T. (2012). Threatening pictures induce shortened time-to-contact estimates. *Attention, perception & psychophysics*, *74*(5), 979–87. doi:10.3758/s13414-012-0285-0

Bresciani, J.-P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of vision*, *6*(5), 554–64. doi:10.1167/6.5.2

Brosch, T., Pourtois, G., & Sander, D. (2010). The perception and categorization of emotional stimuli. *Cognition and Emotion*, *41*, 1–52. doi:10.1080/02699930902975754

Brouwer, A.-M., Van Wouwe, N., Mühl, C., Van Erp, J., & Toet, A. (2013). Perceiving blocks of emotional pictures and sounds: effects on physiological variables. *Frontiers in human neuroscience*, *7*(June), 295. doi:10.3389/fnhum.2013.00295

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.

Byrne, D., Erwin, C. R., & Lamberth, J. (1970). Continuity between the experimental study of attraction and real life computer dating. *Journal of personality and social psychology*, *16*, 157–165.

Calvert, G. A., Campbell, R., & Brammer, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current biology*, *10*(11), 649–57.

Calvert, G. A., Spence, C., & Stein, B. E. (Eds.). (2004). *The Handbook of Multisensory Processes*. Cambridge, MA: MIT Press.

Cannon, W. B. (1929). Bodily Changes in Pain, Hunger, Fear and Rage. *Southern Medical Journal*. doi:10.1097/00007611-192909000-00037

Canzoneri, E., Magosso, E., & Serino, A. (2012). Dynamic sounds capture the boundaries of peripersonal space representation in humans. *PloS one*, *7*(9), e44306. doi:10.1371/journal.pone.0044306

Carlin, A. S., Hoffman, H. G., & Weghorst, S. (1997). Virtual reality and tactile augmentation in the treatment of spider phobia: A case report. *Behaviour Research and Therapy*, *35*(2), 153–158.

Carretié, L., Hinojosa, J. a, López-Martín, S., Albert, J., Tapia, M., & Pozo, M. a. (2009). Danger is worse when it moves: neural and behavioral indices of enhanced attentional capture by dynamic threatening stimuli. *Neuropsychologia*, *47*(2), 364–9. doi:10.1016/j.neuropsychologia.2008.09.007

Chapados, C., & Levitin, D. J. (2008). Cross-modal interactions in the experience of musical performances: physiological correlates. *Cognition*, *108*(3), 639–51. doi:10.1016/j.cognition.2008.05.008

Chen, Y., Edgar, J. C., Holroyd, T., Dammers, J., Thönnessen, H., Roberts, T. P. L., & Mathiak, K. (2010). Neuromagnetic oscillations to emotional faces and prosody. *The European journal of neuroscience*, *31*(10), 1818–27. doi:10.1111/j.1460-9568.2010.07203.x

Choi, Y. H., Jang, D. P., Ku, J. H., Shin, M. B., & Kim, S. I. (2001). Short-term treatment of acrophobia with virtual reality therapy (VRT): a case report. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *4*(3), 349–54.

Clavagnier, S., Falchier, A., & Kennedy, H. (2004). Long-distance feedback projections to area V1: implications for multisensory integration, spatial awareness, and visual consciousness. *Cognitive, affective & behavioral neuroscience*, *4*(2), 117–26.

Coelho, C. M., Santos, J. A., Silvério, J., & Silva, C. F. (2006). Virtual Reality and Acrophobia: One-Year Follow-Up and Case Study. *Cyberpsychology & Behavior*, *9*(3), 336–342.

Coello, Y., Bourgeois, J., & Iachini, T. (2012). Embodied perception of reachable space: How do we manage threatening objects? *Cognitive Processing*, *13*(1 SUPPL).

Colby, C. L., Duhamel, J. R., & Goldberg, M. E. (1993). Ventral intraparietal area of the macaque: anatomic location and visual response properties. *Journal of neurophysiology*, *69*(3), 902–914.

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. doi: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.

Côté, S., & Bouchard, S. (2008). Virtual Reality Exposure for Phobias: A Critical Review. *Journal of CyberTherapy & Rehabilitation*, *1*(1), 75–94.

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., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*(10), 460–467. doi:10.1016/j.tics.2003.08.014

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., Morris, J. S., & Dolan, R. J. (2005). Unconscious fear influences emotional awareness of faces and voices. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(51), 18682–7. doi:10.1073/pnas.0509179102

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.

Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: effects of stimulus onset and intensity on reaction time. *Perception & psychophysics*, *66*(8), 1388–404.

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. doi: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.

Driscoll, I., Hamilton, D. A., Yeo, R. A., Brooks, W. M., & Sutherland, R. J. (2005). Virtual navigation in humans: The impact of age, sex, and hormones on place learning. *Hormones and Behavior*, *47*(3), 326–335.

Duhamel, J.-R., Colby, C. L., & Goldberg, M. E. (1998). Ventral intraparietal area of the macaque: congruent visual and somatic response properties. *Journal of neurophysiology*, *79*(1), 126–36.

Ekman, P., & Friesen, W. V. (1971). Constants across cultures in the face and emotion. *Journal of personality and social psychology*, *17*(2), 124–9.

Emmelkamp, P. M. G., Bruynzeel, M., Drost, L., & Van Der Mast, C. A. P. G. (2001). Virtual reality treatment in acrophobia: a comparison with exposure in vivo. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *4*(3), 335–9.

Emmelkamp, P. M. G., Krijn, M., Hulsbosch, A. M., De Vries, S., Schuemie, M. J., & Van Der Mast, C. A. P. G. (2002). Virtual reality treatment versus exposure in vivo: a comparative evaluation in acrophobia. *Behaviour research and therapy*, *40*(5), 509–16.

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–33. doi:10.1038/415429a

Ernst, M. O., & Bülthoff, H. H. (2004). Merging the senses into a robust percept. *Trends in cognitive sciences*, *8*(4), 162–9. doi:10.1016/j.tics.2004.02.002

Ethofer, T., Anders, S., Erb, M., Droll, C., Royen, L., Saur, R., … Wildgruber, D. (2006). Impact of voice on emotional judgment of faces: an event-related fMRI study. *Human brain mapping*, *27*(9), 707–14. doi:10.1002/hbm.20212

Ethofer, T., Pourtois, G., & Wildgruber, D. (2006). Investigating audiovisual integration of emotional signals in the human brain. *Progress in brain research*, *156*, 345–61. doi:10.1016/S0079-6123(06)56019-4

Fain, G. L. (2003). *Sensory Transduction*. Sunderland, MA: Sinauer Associates.

Falchier, A., Clavagnier, S., Barone, P., & Kennedy, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *The Journal of Neuroscience*, *22*(13), 5749–59. doi:20026562

Farnè, A., & Làdavas, E. (2002). Auditory peripersonal space in humans. *Journal of cognitive neuroscience*, *14*(7), 1030–43. doi:10.1162/089892902320474481

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

Foxe, J. J., Wylie, G. R., Martinez, A., Schroeder, C. E., Javitt, D. C., Guilfoyle, D., … Murray, M. M. (2002). Auditory-Somatosensory Multisensory Processing in Auditory Association Cortex : An fMRI Study. *Journal of Neurophysiology*, *88*, 540–543.

Gagnon, K. T., Geuss, M. N., & Stefanucci, J. K. (2013). Fear influences perceived reaching to targets in audition, but not vision. *Evolution and Human Behavior*, *34*(1), 49–54.

Garcia-Palacios, A., Hoffman, H., Carlin, A., Furness, T. A., & Botella, C. (2002). Virtual reality in the treatment of spider phobia: a controlled study. *Behaviour research and therapy*, *40*(9), 983–93.

Garrido, M. I., Kilner, J. M., Stephan, K. E., & Friston, K. J. (2009). The mismatch negativity: a review of underlying mechanisms. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology*, *120*(3), 453–63. doi:10.1016/j.clinph.2008.11.029

Gentilucci, M., Fogassi, L., Luppino, G., Matelli, M., Camarda, R., & Rizzolatti, G. (1988). Functional organization of inferior area 6 in the macaque monkey - I. Somatotopy and the control of proximal movements. *Experimental Brain Research*, *71*(3), 475–490.

Giard, M. H., & Peronnet, F. (1999). Auditory-visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study. *Journal of cognitive neuroscience*, *11*(5), 473–90.

Gifford, R. (1982). Projected Interpersonal Distance and Orientation Choices: Personality, Sex, and Social Situation. *Social Psychology Quarterly*, *45*(3), 145.

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. doi:10.1016/j.neuropsychologia.2005.09.009

Graziano, M. S. A., & Gross, C. G. (1993). A bimodal map of space: somatosensory receptive fields in the macaque putamen with corresponding visual receptive fields. *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, *97*(1), 96–109.

Graziano, M. S. A., & Gross, C. G. (1995). The representation of extrapersonal space: A possible role for bimodal visuo-tactile neurons. In M. Gazzaniga (Ed.), *The Cognitive Neuroscience* (pp. 1021–1034). Cambridge, MA: MIT Press.

Hagan, C. C., Woods, W., Johnson, S., Calder, A., Green, G. G. R., & Young, A. W. (2009). MEG demonstrates a supra-additive response to facial and vocal emotion in the right superior temporal sulcus. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(47), 20010–5. doi:10.1073/pnas.0905792106

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). doi: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. doi:10.1037//0033-2909.85.1.117

Hayduk, L. A. (1981a). The shape of personal space: An experimental investigation., *13*(1), 87–93. doi:10.1037/h0081114

Hayduk, L. A. (1981b). The permeability of personal space. *Canadian Journal of Behavioural Science/Revue canadienne des sciences du comportement*, *13*(3), 274–287. doi:10.1037/h0081182

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

Helbig, H. B., & Ernst, M. O. (2008). Visual-haptic cue weighting is independent of modality-specific attention. *Journal of vision*, *8*(1), 1–16. doi:10.1167/8.1.21.

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.

Hodges, L. F., Kooper, R., Meyer, T. C., Rothbaum, B. O., Opdyke, D., De Graaff, J. J., … North, M. M. (1995). Virtual environments for treating the fear of heights. *Computer*, *28*(7), 27–33.

Hodges, L. F., Watson, B. A., Kessler, G. D., Rothbaum, B. O., & Opdyke, D. (1996). Virtually conquering fear of flying. *IEEE Computer Graphics and Applications*, *16*(6), 42–49.

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.

Howard, J. P., & Templeton, W. B. (1966). *Human Spatial Orientation*. New York, NY: Wiley.

Iriki, A., Tanaka, M., & Iwamura, Y. (1996). Coding of modified body schema during tool use by macaque postcentral neurones. *Neuroreport*, *7*(14), 2325–2330.

James, W. (1884). What is an Emotion? *Mind*, *9*(34), 188–205. doi:10.1093/mind/os-IX.34.188

Jeffrey, P., & Mark, G. (1998). Constructing social spaces in virtual environments: A study of navigation and interaction. In K. Höök, A. Munro, & D. Benyon (Eds.), *Workshop on Personalised and Social Navigation in Information Space* (pp. 24–38). Stockholm: Swedish Institue of Computer Science.

Jeong, J.-W., Diwadkar, V. A., Chugani, C. D., Sinsoongsud, P., Muzik, O., Behen, M. E., … Chugani, D. C. (2011). Congruence of happy and sad emotion in music and faces modifies cortical audiovisual activation. *NeuroImage*, *54*(4), 2973–82. doi:10.1016/j.neuroimage.2010.11.017

Jessen, S., & Kotz, S. A. (2011). The temporal dynamics of processing emotions from vocal, facial, and bodily expressions. *NeuroImage*, *58*(2), 665–674. doi:10.1016/j.neuroimage.2011.06.035

Jiang, H., Lepore, F., Ptito, M., & Guillemot, J. P. (2004a). Sensory modality distribution in the anterior ectosylvian cortex (AEC) of cats. *Experimental brain research*, *97*(3), 404–14.

Jiang, H., Lepore, F., Ptito, M., & Guillemot, J. P. (2004b). Sensory interactions in the anterior ectosylvian cortex of cats. *Experimental brain research*, *101*(3), 385–96.

Jones, E. G., & Powell, T. P. S. (1970). An anatomical study of converging sensory pathways within the cerebral cortex of the monkey. *Brain*, *91*, 793–820.

Kennedy, D. P., Gläscher, J., Tyszka, J. M., & Adolphs, R. (2009). Personal Space Regulation by the Human Amygdala. *Nature neuroscience*, *12*(10), 1226–1227. doi:10.1038/nn.2381.Personal

Kim, Y. M., Chun, M. H., Yun, G. J., Song, Y. J., & Young, H. E. (2011). The Effect of Virtual Reality Training on Unilateral Spatial Neglect in Stroke Patients. *Annals of Rehabilitation Medicine*. doi:10.5535/arm.2011.35.3.309

Kinchla, R. A. (1974). Detecting target elements in multielement arrays: A confusability model. *Perception & Psychophysics*, *15*(1), 149–158. doi:10.3758/BF03205843

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. doi: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. doi:10.1515/revneuro-2012-0040

Klasen, M., Kenworthy, C. A., Mathiak, K. A., Kircher, T. T. J., & Mathiak, K. (2011). Supramodal representation of emotions. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *31*(38), 13635–43. doi:10.1523/JNEUROSCI.2833-11.2011

Klatzky, R. (1998). Allocentric and egocentric spatial representations: Definitions, distinctions, and interconnections. In C. Freksa, C. Habel, & K. F. Wender (Eds.), *Spatial Cognition - An Interdisciplinary Approach to Representation and Processing of Spatial Knowledge* (pp. 1–17). Heidelberg.

Kober, H., Barett, L., Joseph, J., Bliss-Moreau, E., Lindquist, K., & Wager, T. D. (2008). Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *NeuroImage*, *42*(2), 998–1031. doi:10.1016/j.neuroimage.2008.03.059.Functional

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. doi:10.1016/j.neuroimage.2007.06.020

Krijn, M., Emmelkamp, P. M. G., Biemond, R., De Wilde de Ligny, C., Schuemie, M. J., & Van der Mast, C. A. P. G. (2004). Treatment of acrophobia in virtual reality: the role of immersion and presence. *Behaviour research and therapy*, *42*(2), 229–39. doi:10.1016/S0005-7967(03)00139-6

Làdavas, E., & Farnè, A. (2004). Visuo-tactile representation of near-the-body space. *Journal of Physiology Paris*, *98*(1-3 SPEC. ISS.), 161–170.

Ladavas, E., & Serino, A. (2008). Action-dependent plasticity in peripersonal space representations. *Cognitive neuropsychology*, *25*(7), 1099–113. doi:10.1080/02643290802359113

Lang, P. J., Bradley, M. M., & Cuthbert, B. N. (1997). Motivated attention: Affect, activation, and action. In P. J. Lang, R. F. Simons, & M. T. Balaban (Eds.), *Attention and Orienting: Sensory and Motivational Processes.* (pp. 97–135). Hillsdale, NJ: Lawrence Erlbaum Associates Publishers.

Lang, P. J., & Lazovik, A. D. (1963). Experimental desensitization of a phobia. *Journal of abnormal and social psychology*, *66*(6), 519–25.

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. doi:10.1007/s00221-004-1913-2

Lazarus, R. S. (1991). Cognition and motivation in emotion. *The American psychologist*, *46*(4), 352–367. doi:10.1037/0003-066X.46.4.352

LeDoux, J. (1998). Fear and the brain: where have we been, and where are we going? *Biological psychiatry*, *44*(12), 1229–38.

Lenggenhager, B., Tadi, T., Metzinger, T., & Blanke, O. (2007). Video ergo sum: manipulating bodily self-consciousness. *Science (New York, N.Y.)*, *317*(5841), 1096–1099.

Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Brain research. Cognitive brain research*, *16*(3), 468–78.

Lewis, M., Haviland-Jones, J. M., & Barrett, L. F. (Eds.). (2008). *Handbook of Emotions*. *Handbook of Emotions* (The Guilfo., Vol. 24, pp. 709–730). New York, NY. doi:10.2307/2076468

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*. doi:10.1093/cercor/bht228

Liebowitz, M. R. (1987). Social phobia. *Modern Problems of Pharmacopsychiatry*, *22*, 141–173.

Lloyd, D. M., Coates, A., Knopp, J., Oram, S., & Rowbotham, S. (2009). Don't stand so close to me: The effect of auditory input on interpersonal space. *Perception*, *38*(4), 617–620. doi:10.1068/p6317

Lloyd, D. M., Morrison, I., & Roberts, N. (2006). Role for human posterior parietal cortex in visual processing of aversive objects in peripersonal space. *Journal of neurophysiology*, *95*(1), 205–214.

Lourenco, S. F., & Longo, M. R. (2009). The plasticity of near space: evidence for contraction. *Cognition*, *112*(3), 451–6. doi:10.1016/j.cognition.2009.05.011

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

Makin, T. R., Holmes, N. P., & Zohary, E. (2007). Is that near my hand? Multisensory representation of peripersonal space in human intraparietal sulcus. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, *27*(4), 731–740.

Maravita, A., & Iriki, A. (2004). Tools for the body (schema). *Trends in cognitive sciences*, *8*(2), 79–86. doi:10.1016/j.tics.2003.12.008

Massaro, D., & Egan, P. (1996). Perceiving affect from the voice and the face. *Psychonomic Bulletin & Review*, *3*(2), 215–221.

Meredith, M. A., Nemitz, J. W., & Stein, B. E. (1987). Determinants of Multisensory Integration Neurons. *The Journal of Neuroscience*, *7*(10), 3215–3229.

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. doi:10.1093/cercor/bhh007

Mühlberger, A., Herrmann, M. J., Wiedemann, G. C., Ellgring, H., & Pauli, P. (2001). Repeated exposure of flight phobics to flights in virtual reality. *Behaviour research and therapy*, *39*(9), 1033–50.

Mühlberger, A., Sperber, M., Wieser, M. J., & Pauli, P. (2008). A Virtual Reality Behavior Avoidance Test (VR-BAT) for the assessment of spider phobia. *Journal of CyberTherapy & Rehabilitation*, *1*(2), 147–158.

Müller, V. I., Habel, U., Derntl, B., Schneider, F., Zilles, K., Turetsky, B. I., & Eickhoff, S. B. (2011). Incongruence effects in crossmodal emotional integration. *NeuroImage*, *54*(3), 2257–66. doi:10.1016/j.neuroimage.2010.10.047

Murray, M. M., Foxe, J. J., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2001). Visuo-spatial neural response interactions in early cortical processing during a simple reaction time task: a high-density electrical mapping study. *Neuropsychologia*, *39*(8), 828–44.

Murray, M. M., Molholm, S., Michel, C. M., Heslenfeld, D. J., Ritter, W., Javitt, D. C., … Foxe, J. J. (2005). Grabbing your ear: rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment. *Cerebral cortex*, *15*(7), 963–74. doi:10.1093/cercor/bhh197

Newell, F. N., Ernst, M. O., Tjan, B. S., & Bülthoff, H. H. (2001). Viewpoint dependence in visual and haptic object recognition. *Psychological science*, *12*, 37–42.

Ngo, M. K., & Spence, C. (2010). Auditory , tactile , and multisensory cues facilitate search for dynamic visual stimuli. *Attention, perception & psychophysics*, *72*(6), 1654–1665. doi:10.3758/APP

North, M. M., North, S. M., & Coble, J. R. (1997). Virtual reality therapy for fear of flying. *American Journal of Psychiatry*.

North, M. M., North, S. M., & Coble, J. R. (1998). Virtual reality therapy: an effective treatment for phobias. *Studies in health technology and informatics*, *58*, 112–119.

O'Neal, E. C., Brunault, M. A., Marquis, J., & Carifio, M. S. (1979). Anger and the body-buffer zone. *Journal of social psychology*, *108*, 135–136.

Olofsson, J. K., Nordin, S., Sequeira, H., & Polich, J. (2008). Affective picture processing: An integrative review of ERP findings. *Biological psychologia*, *77*(3), 247–265.

Park, J.-Y., Gu, B.-M., Kang, D.-H., Shin, Y.-W., Choi, C.-H., Lee, J.-M., & Kwon, J. S. (2010). Integration of cross-modal emotional information in the human brain: an fMRI study. *Cortex; a journal devoted to the study of the nervous system and behavior*, *46*(2), 161–9. doi:10.1016/j.cortex.2008.06.008

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. doi:10.1016/S0006-3223(03)00168-9

Pick, H. L., Warren, D. H., & Hay, J. C. (1969). Sensory conflict in judgments of spatial direction. *Perception And Psychophysics*, *6*(4), 203–205.

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.

Pourtois, G., De Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. *Neuroreport*, *11*(6), 1329–33.

Pourtois, G., Debatisse, D., Despland, P.-A., & De Gelder, B. (2002). Facial expressions modulate the time course of long latency auditory brain potentials. *Brain research. Cognitive brain research*, *14*(1), 99–105.

Raab, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions Of The New York Academy Of Sciences*, *24*, 574–590.

Reinoso-Suarez, F., & Roda, J. M. (1985). Topographical organization of the cortical afferent connections to the cortex of the anterior ectosylvian sulcus in the cat. *Experimental brain research*, *59*, 313–324.

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

Riva, G. (2005). Virtual reality in psychotherapy: review. *Cyberpsychology & behavior : the impact of the Internet, multimedia and virtual reality on behavior and society*, *8*(3), 220–30; discussion 231–40. doi:10.1089/cpb.2005.8.220

Riva, G., Bacchetta, M., Baruffi, M., Rinaldi, S., & Molinari, E. (1999). Virtual reality based experiential cognitive treatment of anorexia nervosa. *Journal of behavior therapy and experimental psychiatry*, *30*(3), 221–30.

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

Rizzolatti, G., Scandolara, C., Matelli, M., & Gentilucci, M. (1981). Afferent properties of periarcuate neurons in macaque monkeys. *Behavioural brain research*, *2*(2), 125–46.

Robins, D. L., Hunyadi, E., & Schultz, R. T. (2009). Superior temporal activation in response to dynamic audio-visual emotional cues. *Brain and cognition*, *69*(2), 269–78. doi:10.1016/j.bandc.2008.08.007

Rockland, K. S., & Ojima, H. (2003). Multisensory convergence in calcarine visual areas in macaque monkey. *International Journal of Psychophysiology*, *50*(1-2), 19–26. doi:10.1016/S0167-8760(03)00121-1

Rothbaum, B. O., Anderson, P., Zimand, E., Hodges, L. F., Lang, D., & Wilson, J. (2006). Virtual Reality Exposure Therapy and Standard (in Vivo) Exposure Therapy in the Treatment of Fear of Flying. *Behavior Therapy*, *37*(1), 80–90.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M. M. (1995a). Effectiveness of computer-generated (virtual reality) graded exposure in the treatment of acrophobia. *American Journal of Psychiatry*, *152*(4), 626–628.

Rothbaum, B. O., Hodges, L. F., Kooper, R., Opdyke, D., Williford, J. S., & North, M. M. (1995b). Virtual Reality Graded Exposure in the Treatment of Acrophobia : A Case Report. *Behavior therapy*, *26*, 547–554.

Russell, J. A. (1980). A circumplex model of affect. *Journal of personality and social psychology*, *39*(6), 1161–1178.

Saladin, M. E., Brady, K. T., Graap, K., & Rothbaum, B. O. (2006). A preliminary report on the use of virtual reality technology to elicit craving and cue reactivity in cocaine dependent individuals. *Addictive Behaviors*, *31*(10), 1881–1894.

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

Schiffenbauer, A., & Steven Schiavo, R. (1976). Physical distance and attraction: An intensification effect. *Journal of Experimental Social Psychology*, *12*(3), 274–282.

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

Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*.

Senkowski, D., Saint-Amour, D., Höfle, M., & Foxe, J. J. (2011). Multisensory interactions in early evoked brain activity follow the principle of inverse effectiveness. *NeuroImage*, *56*(4), 2200–8. doi:10.1016/j.neuroimage.2011.03.075

Sereno, M. I., & Huang, R.-S. (2006). A human parietal face area contains aligned head-centered visual and tactile maps. *Nature neuroscience*, *9*(10), 1337–1343.

Serino, A., Bassolino, M., Farnè, A., & Làdavas, E. (2007). Extended multisensory space in blind cane users. *Psychological science*, *18*(7), 642–8. doi:10.1111/j.1467-9280.2007.01952.x

Serino, A., Canzoneri, E., & Avenanti, A. (2011). Fronto-parietal areas necessary for a multisensory representation of peripersonal space in humans: an rTMS study. *Journal of cognitive neuroscience*, *23*(10), 2956–67. doi:10.1162/jocn_a_00006

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. doi:10.1111/j.1467-9280.2008.02105.x

Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Brain research. Cognitive brain research*, *14*(1), 147–52.

Slater, M., Spanlang, B., Sanchez-Vives, M. V., & Blanke, O. (2010). First person experience of body transfer in virtual reality. *PLoS ONE*, *5*(5).

Sommer, R. (1959). Studies in personal space. *Sociometry*, *22*(3), 247–260. doi:10.2307/2785668

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.

Spence, C., Pavani, F., Maravita, A., & Holmes, N. (2004). Multisensory contributions to the 3-D representation of visuotactile peripersonal space in humans: Evidence from the crossmodal congruency task. *Journal of Physiology Paris*.

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.

Spreckelmeyer, K. N., Kutas, M., Urbach, T. P., Altenmüller, E., & Münte, T. F. (2006). Combined perception of emotion in pictures and musical sounds. *Brain research*, *1070*(1), 160–70. doi:10.1016/j.brainres.2005.11.075

Stanford, T. R., Quessy, S., & Stein, B. E. (2005). Evaluating the Operations Underlying Multisensory Integration in the Cat Superior Colliculus. *The Journal of Neuroscience*, *25*(28), 6499–6508.

Stein, B. E., & Meredith, M. A. (1993). *The Merging of the Senses*. *Book* (Vol. onpp). Cambridge, MA: MIT 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. doi:10.1038/nrn2331

Stevenson, R. A., & James, T. W. (2009). Audiovisual integration in human superior temporal sulcus: Inverse effectiveness and the neural processing of speech and object recognition. *NeuroImage*, *44*(3), 1210–23. doi:10.1016/j.neuroimage.2008.09.034

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. doi: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. doi:10.1163/22134808-00002424

Taffou, M., & Viaud-Delmon, I. (2014). Cynophobic Fear Adaptively Extends Peri-Personal Space. *Frontiers in Psychiatry*, *5*(September), 3–9. doi: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. doi:10.1016/j.neuropsychologia.2008.07.025

Tajadura-Jiménez, A., Pantelidou, G., Rebacz, P., Västfjäll, D., & Tsakiris, M. (2011). I-space: the effects of emotional valence and source of music on interpersonal distance. *PloS one*, *6*(10), e26083. doi:10.1371/journal.pone.0026083

Tajadura-Jiménez, A., Väljamäe, A., Asutay, E., & Västfjäll, D. (2010). Embodied auditory perception: the emotional impact of approaching and receding sound sources. *Emotion*, *10*(2), 216–229. doi:10.1037/a0018422

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. doi:10.1177/0956797610380698

Teneggi, C., Canzoneri, E., Di Pellegrino, G., & Serino, A. (2013). Social Modulation of Peripersonal Space Boundaries. *Current biology : CB*, 1–6. doi:10.1016/j.cub.2013.01.043

Teunisse, J. P., & De Gelder, B. (2001). Impaired categorical perception of facial expressions in high-functioning adolescents with autism. *Child neuropsychology : a journal on normal and abnormal development in childhood and adolescence*, *7*(1), 1–14. doi:10.1076/chin.7.1.1.3150

Thyer, B. A., Papsdorf, J. D., Davis, R., & Vallecorsa, S. (1984). Autonomic correlates of the subjective anxiety scale. *Journal of behavior therapy and experimental psychiatry*, *15*(1), 3–7.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive psychology*, *12*, 97–136.

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

Valdés-Conroy, B., Román, F. J., Hinojosa, J. A., & Shorkey, S. P. (2012). So Far So Good: Emotion in the Peripersonal/Extrapersonal Space. (A. Serino, Ed.)*PLoS ONE*, *7*(11), e49162. doi:10.1371/journal.pone.0049162

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

Van den Stock, J., Peretz, I., Grèzes, J., & De Gelder, B. (2009). Instrumental music influences recognition of emotional body language. *Brain topography*, *21*(3-4), 216–20. doi:10.1007/s10548-009-0099-0

Van den Stock, J., Righart, R., & De Gelder, B. (2007). Body expressions influence recognition of emotions in the face and voice. *Emotion (Washington, D.C.)*, *7*(3), 487–94. doi:10.1037/1528-3542.7.3.487

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. doi:10.1016/j.eurpsy.2004.10.004

Viaud-Delmon, I., Znaïdi, F., Bonneel, N., Doukhan, D., Suied, C., Warusfel, O., … Drettakis, G. (2008). Auditory-visual virtual environments to treat dog phobia. In *The Seventh International Conference on Disability, Virtual Reality and Associated Technologies with ArtAbilitation 2008* (pp. 119–124). Porto.

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. doi: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. doi:10.1016/j.cognition.2005.09.003

Vroomen, J., Bertelson, P., & De Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & psychophysics*, *63*(4), 651–9.

Vroomen, J., Driver, J., & De Gelder, B. (2001). Is cross-modal integration of emotional expressions independent of attentional resources? *Cognitive, affective & behavioral neuroscience*, *1*(4), 382–7.

Wald, J. (2004). Efficacy of virtual reality exposure therapy for driving phobia: A multiple baseline across-subjects design. *Behavior Therapy*, *35*(3), 621–635. doi:10.1016/S0005-7894(04)80035-2

Wald, J., & Taylor, S. (2001). Efficacy of virtual reality exposure therapy to treat driving phobia: a case report. *Journal of behavior therapy and experimental psychiatry*, *31*(3-4), 249–57.

Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological bulletin*, *88*(3), 638–667.

Wiederhold, B. K., Gevirtz, R., & Wiederhold, M. D. (1998). Fear of Flying : A Case Report Using Virtual Reality Therapy with Physiological Monitoring. *Cyberpsychology & Behavior*, *1*(2), 97–104.

Wilcox, L. M., Allison, R. S., Elfassy, S., & Grelik, C. (2006). Personal space in virtual reality. *ACM Transactions on Applied Perception*, *3*(4), 412–428. doi:10.1145/1190036.1190041

Williams, L. E., & Bargh, J. A. (2008). Keeping One's Distance. *Psychological Science*, *19*(3), 302–309.

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. doi:10.1016/j.neuropsychologia.2006.12.004

Résumé :

Induire un ressenti de peur avec la réalité virtuelle
Etude de l'influence de stimuli multisensoriels sur l'expérience émotionnelle négative

Dans l'environnement naturel, les signaux émotionnels sont transmis *via* différentes modalités sensorielles. Par exemple, un chien agressif émet des signaux ayant un impact émotionnel à la fois *via* la modalité visuelle (crocs) et *via* la modalité auditive (grognements ou aboiements). Pourtant, l'effet d'évènements affectifs multisensoriels sur l'expérience émotionnelle consciente (le ressenti) reste relativement peu connu. Est-ce que les stimuli affectifs multisensoriels augmentent le ressenti émotionnel? Le travail de recherche présenté dans cette thèse a exploité les avantages des techniques de réalité virtuelle pour étudier l'expérience émotionnelle négative induite par des évènements aversifs visuo-auditifs présentés dans un contexte écologique. Un tel contexte permet de prendre en compte la distance entre le sujet et le stimulus affectif, qui représente un facteur important puisque les évènements situés près du corps sont représentés différemment des évènements situés loin du corps au niveau cérébral. Par conséquent, ce travail de recherche a impliqué l'étude des liens entre l'affect, la présentation multisensorielle et l'espace.

Une première étude utilisant la réalité virtuelle a testé l'influence de stimuli aversifs visuo-auditifs sur le ressenti. Lors de cette étude, deux groupes de participants (sensibles ou non à la peur des chiens) ont été exposés à des environnements virtuels visuo-auditifs contenant des chiens virtuels. Les participants ont exploré ces environnements virtuels et rapporté l'intensité de leur ressenti de peur en réponse à des chiens virtuels dont la présentation sensorielle pouvait être uniquement visuelle, uniquement auditive, ou visuo-auditive. Les deux groupes de participants, sensibles et non sensibles à la peur des chiens, ont rapporté un ressenti de peur plus intense en réponse aux stimuli bimodaux visuo-auditifs comparés aux stimuli unimodaux. Les résultats de cette étude suggèrent que la présentation multisensorielle de stimuli aversifs amplifie l'expérience émotionnelle négative.

Une deuxième étude a examiné l'effet de la peur excessive sur la représentation de l'espace. La taille de l'espace péri-personnel (proche du corps) de participants sensibles ou non à la phobie des chiens a été mesurée, grâce à une tâche audio-tactile, alors qu'ils entendaient un son de chien ou de mouton qui s'approchait d'eux. Les résultats ont montré que, en présence du son de mouton, la taille de l'espace péri-personnel des participants sensibles à la phobie

des chiens était similaire à celle de l'espace péri-personnel des participants non sensibles à la phobie des chiens. Par contre, en présence du son de chien, l'espace péri-personnel des participants ayant une peur excessive des chiens s'agrandissait. Cet effet de la présence du son de chien n'a pas été retrouvé chez les participants qui ne présentaient pas de peur excessive des chiens. Cette étude a démontré que la sensibilité à la phobie des chiens a une influence sur la représentation de l'espace proche du corps et suggère que l'apparition d'un objet phobogène aux abords de l'espace péri-personnel provoque une extension de la surface de celui-ci.

Une troisième étude réalisée en réalité virtuelle a examiné l'effet de stimuli aversifs visuo-auditifs sur le ressenti en fonction de leur position plus ou moins proche du sujet en utilisant la réalité virtuelle. Lors de cette étude, deux groupes de participants (sensibles ou non à la peur des foules) ont été exposés à un environnement virtuel visuo-auditif contenant des foules virtuelles. Les participants ont exploré cet environnement virtuel et rapporté l'intensité de leur inconfort quand ils se trouvaient à une distance lointaine ou proche de foules virtuelles dont la présentation sensorielle pouvait être uniquement visuelle, uniquement auditive, ou visuo-auditive. Les participants sensibles à la peur des foules ont rapporté un ressenti plus intense en réponse aux foules bimodales visuo-auditives qu'en réponse aux foules unimodales. Cet effet n'a été observé que lorsque les foules se trouvaient à une distance proche des participants. Quand les foules se trouvaient à une distance plus lointaine, la présentation sensorielle n'a pas eu d'influence sur l'intensité du ressenti rapporté par les participants. Les résultats de cette troisième étude sont cohérents avec les résultats de la première étude car ils confirment que la présentation multisensorielle de stimuli aversifs amplifie l'expérience émotionnelle négative. De plus, ils suggèrent que l'effet de la présentation sensorielle sur l'expérience émotionnelle négative dépend de la localisation spatiale des stimuli aversifs.

En conclusion, il a été constaté que le ressenti émotionnel est modulé par les caractéristiques sensorielles et spatiales des évènements aversifs. Les stimuli aversifs visuo-auditifs amplifient le ressenti négatif. Cependant, cet effet n'existe que si ces stimuli sont dans l'espace proche du sujet. Enfin, la peur excessive d'un stimulus spécifique provoque une extension de l'espace péri-personnel. Il semble donc important d'évaluer la taille de l'espace péri-personnel des sujets afin de pouvoir contrôler la position du stimulus aversif dans l'espace proche ou lointain lors de l'étude de l'expérience émotionnelle induite par des stimuli multisensoriels.

L’ensemble de ces travaux fournit de nouvelles informations sur le traitement de l’information affective et suggère que les caractéristiques sensorielles et spatiales des stimuli affectifs sont des variables importantes à prendre en compte dans l’étude de l’affect chez l’homme. Ces travaux mettent également en évidence l'utilité et la pertinence de la réalité virtuelle pour l'étude de l'affect. En effet, la réalité virtuelle peut aider à mieux comprendre l’affect car elle permet une présentation plus écologique des stimuli affectifs et facilite l’étude de leurs aspects sensoriel et spatial. De plus, les résultats sont directement exploitables pour les thérapies en réalité virtuelle et peuvent aider à affiner le développement d'environnements virtuels pour le traitement de troubles émotionnels.

Mots clés : intégration visuo-auditive, expérience émotionnelle consciente, peur, espace péri-personnel, réalité virtuelle, thérapie d’exposition en réalité virtuelle

Abstract :

In a natural environment, affective events often convey emotional cues through multiple sensory modalities: the aggressiveness of a dog has both visual and auditory manifestations. Yet, the effect of multisensory affective events on the conscious emotional experience (feelings) they induce remains relatively undiscovered. The research presented in this thesis exploited the unique advantages of virtual reality techniques to examine the negative emotional experience induced by auditory-visual aversive events embedded in a natural context. In natural contexts, the spatial distance between the perceiver and the affective stimuli is an important factor, given that events located at close or far distances are represented differently in the brain. Consequently, the present research included the investigation of the relationship between affect, multisensory presentation and space.

A first study using virtual reality tested the influence of auditory-visual aversive stimuli on negative emotional experience. A second study explored the effect of excessive fear on the representation of close space. A third study examined the effect of auditory-visual stimuli on negative emotional experience as a function of their location at close or far distances from the perceiver.

Overall, it was found that negative emotional experience is modulated by the sensory and spatial characteristics of aversive events. Multisensory aversive events amplify negative feelings only when they are located at close distances from the perceiver. Moreover, excessive fear related to an event extends the space, wherein the event is represented as close.

Taken together, the present research provides new information about affective processing and suggests that sensory and spatial factors are important variables to take into account in the investigation of affect. It also exposes virtual reality as a relevant tool for the study of human affect. Virtual reality might help us to better understand affective processing by providing more ecological stimulation and thus allowing for the investigation of factors seemingly involved in everyday human affective experience, such as spatial and sensory factors. Moreover, the findings can be directly exploited in research on virtual reality-based therapy and help developing refined virtual environments for the treatment of emotional disorders.

Keywords: auditory-visual integration, conscious emotional experience, fear, near space, virtual reality, virtual reality exposure therapy