Auditory-visual virtual environments to treat dog phobia

Isabelle Viaud-Delmon, Feryel Znaïdi, Nicolas Bonneel, Clara Suied, Olivier Warusfel, N'Guyen Khoa-Van, George Drettakis

AUDITORY-VISUAL VIRTUAL ENVIRONMENTS TO TREAT DOG PHOBIA

I Viaud-Delmon[1], F Znaïdi[2], N Bonneel[3], D Doukhan[4], C Suied[5], O Warusfel[6], K V N'Guyen[7] and G Drettakis[8]

[1, 2, 5]CNRS UPMC UMR 7593, La Salpetriere Hospital & IRCAM, 1 place Igor Stravinsky, Paris, FRANCE

[3, 8]REVES INRIA, Sophia Antipolis, FRANCE

[4, 6, 7]IRCAM, 1 place Igor Stravinsky, Paris, FRANCE

[1]*ivd@ext.jussieu.fr*, [2]*feryelz@gmail.com*, [3]*nicolas.bonneel@sophia.inria.fr*, [4]*david.doukhan@ircam.fr,* [5]*clara.suied@ircam.fr,* [6]*olivier.warusfel@ircam.fr,* [7]*khoa-van.nguyen@ircam.fr,* [8]*george.drettakis@sophia.inria.fr*

[1,2,4-7]*www.ircam.fr,* [3,8]*www-sop.inria.fr/reves*

ABSTRACT

In this paper we present the design, development, and usability testing of an auditory-visual based interactive environment for investigating virtual reality exposure-based treatment for cynophobia. The application is developed upon a framework that integrates different algorithms of the CROSSMOD project (www.crossmod.org). We discuss the on-going work and preliminary observations, so as to further the development of auditory-visual environment for virtual reality. Traditionally, virtual reality concentrates primarily on the presentation of high fidelity visual experience. We aim at demonstrating that combining adequately the visual and the auditory experience provides a powerful tool to enhance sensory processing and modulate attention.

1. INTRODUCTION

Virtual Reality (VR) has been employed as an alternative for in vivo exposure for the treatment of different phobias for the past decade (see Riva, 2005; Parsons and Rizzo, 2008; Cobb and Sharkey, 2006). Specific phobias have been successfully treated with VR, including arachnophobia (e.g. Carlin et al, 1997). However, few of these procedures have been conducted with VR involving multiple sensory stimulations. Modulation of perceptual processing depends on sensory inputs from all modalities, and can potentially influence perception and behaviour in multiple ways. The aim of this study is to investigate VR exposure-based treatment for cynophobia (dog phobia). We have chosen to work on cynophobia because the acoustic aspect of this phobia is much more relevant than in some other phobias, and is therefore an ideal target to test the efficacy of auditory and visual environments to generate presence and emotion. Exposure-based treatments for cynophobia can use in vivo techniques, imaginal techniques, or hybrid technique (Rentz et al, 2003). In order to treat a phobia, the virtual exposure needs to evoke genuine reactions and emotions. In the case of cynophobia, a very important component of the anxiogenic stimulus is the auditory one: dogs barking and growling are very efficient in provoking emotional reactions in any individual.

Diverse audio-based applications have been implemented in the last few years, involving 3D sound. However, the majority of these applications have been designed to work with blind persons (e.g. Lahav and Mioduser, 2002; Sanchez, 2004) or individuals with severe disability (Brooks et al, 2002). In our application, auditory information is not used as a way to supplement visual information. Because strong emotional reactions can easily be elicited through audition (Bremner et al, 1999), we want to fully exploit the potentiality of 3D audio to increase the realism and richness of the immersive environment. Humans are easily distracted by the locus of an irrelevant auditory event when their visual attention is focused elsewhere

(Van der Lubbe and Postma, 2005). This implies that endogenous orienting does not suppress auditory exogenous effects, and that audition can therefore serve as a strong cue to induce emotional reactions in individuals immersed in a visual environment. Furthermore, animal sounds seem to have a stronger influence in alerting than any other sounds (Suied and Viaud-Delmon, 2008).

Therefore, the study we present here involves technologies, models and applications linked to the use of 3D sound in VR environments. Auditory augmentation of visual environments is known to improve presence and immersion (Hendrix and Barfield, 1996). To create such environments and the corresponding content, several concepts and technologies need to be researched, developed, and/or integrated. The introduction of 3D sound also addresses the need for a better understanding of multisensory integration mechanisms. This includes complementarities or conflict between the auditory and visual senses and also with idiothetic cues (cues generated through self-motion, including vestibular and proprioceptive information).

The most natural audio technique for VR applications is the binaural rendering on headphones that relies on the use of HRTFs. HRTF refers to Head Related Transfer Function, which is a set of filters measured on an individual or artificial head and used to reproduce all the directional cues involved in auditory localization (Blauert, 1983) However, incorporating real-time updated 3D sound to VR technologies addresses several practical issues. If there is a consensus on the fact that presence is improved by 3D sound, little is known about how an auditory VE should be designed so that it does not interfere with the visual VE (Vastfjall, Larsson and Kleiner, 2002). To develop an application planned for clinical use, we therefore have to conduct several preliminary studies to know how to design an auditory environment and how to integrate efficiently the auditory component of the threatening stimulus.

The primary aim of the current study is to determine the situations in which emotional reactions can be evoked in individuals who fear dogs. A secondary aim is to test the efficacy of progressive exposure thanks to features that can be manipulated in VR only (e.g. visual contrast of the scene, presence or absence of visuals when there is a sound, coherence or conflict of the sound with the visuals, sound source localisation control, dog behavioural control etc…). We will present here all the necessary steps which guided us to choose the different components of the virtual environments, in which participants will be immersed and confronted to virtual dogs. Our final goal will be to determine whether dog phobia can be successfully ameliorated with virtual reality treatment.

Exposure techniques have to be progressive (e.g., the first time you only see a dog far away, second time it's a bit closer, third time it is next to you, fourth time you are surrounded by dogs). We have chosen to develop gradation techniques, which vary along several dimensions to manipulate the emotional component of the stimulus. In this study, we use 3D sound as a progressive exposure tool. We use different sounds of dogs, accompanied or not with visuals, graded in emotional valence through the manipulation of the light and the composition of the visual environment.

2. METHODS

2.1 Procedure

To conduct our study, we needed to select participants reporting an unusual fear of dogs. We also needed to gather information on fear of dogs to design our virtual environment and animated dogs. We therefore conducted a two-stage screening process. On the basis of this screening, we invited 10 participants to take part to the evaluation procedure of our application and selected one dog model.

2.1.1 Fear of dogs screening

In order to select our participants, we developed a preliminary questionnaire assessing fear of dogs (possible range 0-42). This questionnaire was composed of 14 items 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). Four preliminary yes/no questions were asked: "Do you have fear of dogs?", "Do you avoid dogs or endure their presence with intense anxiety?", "Are you afraid of a specific dog breed and if yes, which one?", "Does the size of dog have an effect on your fear?" Seventy-five individuals (32 females) have participated to this screening. Mean age of the sample was 31.3 years (SD=8.4).

2.1.2 Virtual dogs rating

In order to validate the threatening dog model on which our exposure protocol would be based, we first built 9 different dog models and animated them with a growling movement. We used the following dogs breed:

boxer, German shepherd, pit-bull, Staffordshire, Doberman, Miniature Pinscher, Malamute, bull terrier, Great Dane. We then asked 10 participants to rate the valence and arousal of animations of these different dog models. On the computer screen, each of the 9 dogs was presented for 1 s. At picture offset, the self-assessment rulers for valence and arousal were presented. Presentation of dog models was pseudo-randomised.

2.2 Virtual environments

We use exterior and interior virtual scenes. The two exterior scenes are a street scene (see Figure 1), with cars passing by and traffic noise; and a garden scene with trees, a house, tables and benches. The interior scene is a large dark hangar, in which different industrial machinery are active and producing contact sounds.

Figure 1. *The three virtual environments used in this study. On the left, the city scene, in the middle, the garden scene, and on the right, the factory scene.*

2.3 Setup

The sessions take place in an acoustically damped and sound proof recording studio with the light switched off. The visual scenes are presented on a 300×225 cm² stereoscopic passive screen, corresponding to 90×74 degs at the viewing distance of 1.5 m, and are projected with two F2 SXGA+ ProjectionDesign projectors. Participants wear polarized glasses.

Figure 2. *Virtual Reality setup. On the left, view of the entire screen in the acoustically damped studio, on the right, a participants equipped with headphones, a head tracker and polarized glasses.*

The auditory scenes are presented through Sennheiser HD650 headphones. To provide good virtual auditory localization cues, the sound stimuli are processed through binaural rendering (Blauert 1983), using selected non-individual HRTF of the LISTEN HRTF database (http://recherche.ircam.fr/equipes/salles/listen/). Head movements are tracked so that stereo and 3D sounds are appropriately rendered. The participants are equipped with a wireless mouse to navigate in the virtual environment. With this device, they control both rotations and translations within the virtual scene.

2.4 Rendering

Virtual three-dimensional visual scenes are rendered with the OgreVR API developed by INRIA. OgreVR is the combination between the graphics rendering engine Ogre (Milne and Rowe, 2002) and the library VRPN (Hudson et al 2001), which provides a transparent interface between an application program and most of the devices used in virtual reality such as tracking systems, flysticks, joysticks etc. These 2 libraries are open source projects written in C++. The rendering machine is based on an AMD Dual Core Opteron 275 (2.2 GHz). The machine is equipped with two 2 GeForce 7950 GX2 M500 graphic cards.

2.5 Evaluation of the application

Evaluation is made according to the capacity of the virtual environments to evoke fear in participants. The final evaluation will be made with the presence scores, the Subjective Unit of Distress scores (Wolpe, 1973), and the level of improvement in fear.

2.5.1 Questionnaires and Interview Measures related to the exposure sessions

The state portion of the STAI (Spielberger et al, 1983) is used to measure the anxiety levels upon arrival at the laboratory and after completion of the exposure session. A 22-item cybersickness scale assesses the level of discomfort after exposure to VR (Viaud-Delmon et al, 2000). It 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 presence questionnaire from the I-group (Schubert et al 2001) is presented after immersion.

Anxiety ratings are collected during the exposure with Subjective Unit of Distress (Wolpe, 1973). The SUDs rating is a valid self report measurement of anxiety on a 0–100 point scale. Scores of 0 represent no fear and 100 represents the most fear the individual has ever felt in their life. SUDs ratings are taken in 5 min intervals throughout the session.

2.5.2 Measures during the exposure sessions

Participants have to fulfil a task during the exposure. In each of the environment, they have to find targets and to validate them in order to go on with the exploration. In order to find these targets, they need to follow a trajectory in which they will encounter dogs.

We therefore measure the success to each task, but also count the behavioural reactions of the participants whenever they encounter a dog (step backward, freezing …).

3. RESULTS AND DESCRIPTION

3.1 Fear of dogs screening

The preliminary questionnaire obtained a mean rating of 10.4 (SD=9). There was no difference between males and females ratings. Fourteen individuals responded “yes” to the first question (“Do you have fear of dogs?”). These individuals had scores ranging from 8 to 34 (mean=23.3, SD=9.3). Therefore, this yes/no question was not completely discriminative and the total score might provide more information.

Thirty one individuals reported that they are afraid of a specific breed. Pit-bull were reported 14 times, Doberman 7 times, then Rottweiler (4), German Shepard (3) and Bulldog (3). Forty-four individuals reported that the size of a dog had an impact on their emotional reaction.

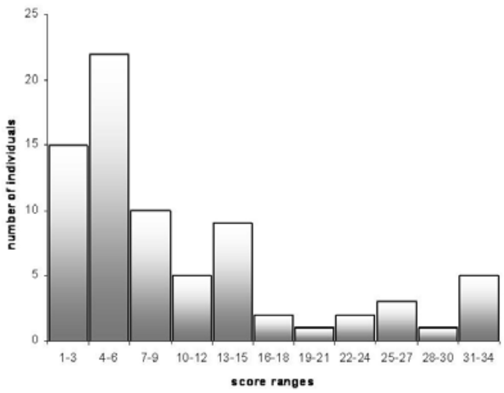

Figure 3. *Scores distribution of fear of dogs in a sample of 75 individuals.*

Participants who obtained a total score superior to 20 (values above 1 standard deviation of the mean of the total sample) were invited to take part in a diagnostic interview with a clinical psychologist. Selected individuals were then invited to further participate in exposure sessions, aiming at reducing their fear of dogs. Ten participants are currently taking part in the study.

3. 2 Virtual dog selection

Ten participants in the fear of dogs screening were further invited to evaluate 9 animated dogs, modelled after the answers gathered in the questionnaire. Surprisingly, the pit-bull model was not judged as the most negatively valenced and arousing, and the Doberman model was the most uniformly rated.

The Doberman was therefore selected as the threatening stimulus for the exposure sessions. Several animations have been developed: running, walking, seating, jumping etc. It also affords growling as well as barking, and the experimenter can control the dog animations with keys. The dog's barking and growling are spatialised in 3D.

Figure 4. *The Doberman dog model, which was selected as the dog to be included in the exposure sessions.*

3. 3 Pre-evaluation of the application: the scenarii

Participants who have been exposed to the on-going work have provided valuable comments to improve the application. The first scene to which they were exposed was the city scene. In this city, they had to find a restaurant, then to go to a cash distributor machine, after which they had to follow a dog to find the endpoint of their journey. This scene was evaluated both with a dim light with sharp shadows and a bright light. The auditory environment was composed of cars roaring, and sounds recorded in a city. Subjects met a silent dog when they walked towards the restaurant. The next dog was seating next to the cash distributor machine, growling as they were approaching. The dog they had to follow was not emitting any sound. Participants did not notice any difference across the two light conditions, and completely focused their attention on the task. Obviously, perceptual processing was enhanced by the emotional cues: the dogs in this scene were presented with a low contrast. Still, participants did not notice any difference with the scene in which the dogs were presented with a high contrast.

In the second scene, participants had to find a specific house, while walking in front of several houses surrounded by fences. Dogs were barking behind these fences. The auditory environment was composed of different outdoor sounds, including birds. While the visual environment was judged as less anxiogenic than the city, the dogs barking did elicit a high level of anxiety.

In the last scene, participants had to walk across an automatic door guarded by a dog, which would stand up and growl as they were approaching. The scene was composed of several sounds emitted by the machinery. This was judged as the most anxiogenic and the task as extremely difficult.

This pre-evaluation indicates that sensory processing and attention was modulated by the affective significance of stimuli. Participants did pay more attention to emotional stimuli, as expected. Even in a dim light, emotional cues enhanced visual processing of the stimuli.

4. CONCLUSIONS

The on-going evaluation of our application aims at testing the capacity of the virtual environments to evoke fear in participants. The final evaluation will be made with the presence scores, the SUD scores, and the level of improvement in fear. First results are very encouraging and indicate that strong emotional reactions can be elicited and mastered through a semantically gradual exposure. The preliminary evaluation showed that the environments did produce an increased perceptual processing of emotional stimuli, resulting in impairment in doing the proposed tasks. Apart from spatial orienting effects, auditory stimuli are known to increase alertness. Our application provide evidence that these effects may also be modulated by endogenous orienting, and, the other way round, that it may be more difficult to keep attention focused in the case of highly alerting stimuli. When taken into account, these specific properties of auditory stimuli can be efficiently integrated in a virtual environment to influence endogenous orienting.

Acknowledgements: This research is supported by the EU IST FP6 Open FET project CROSSMOD.

5. REFERENCES

J Blauert (1983), *Spatial hearing*, MIT Press, Cambridge.

J D Bremner, L H Staib, D Kaloupek, S M Southwick, R Soufer and D S Charney (1999), Neural correlates of exposure to traumatic pictures and sound in Vietnam combat veterans with and without posttraumatic stress disorder: a positron emission tomography study, *Biol Psychiatry*, **45**, 806–816.

T Brooks, A Camurri, N Canagarajah and S Hasselblad (2002), Interaction with shapes and sounds as a therapy for special needs and rehabilitation, *Proc. 4th Intl. Conf. Disability, Virtual Reality & Assoc. Tech.,* Veszprem, Hungary, pp. 205–212.

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

S V G Cobb and P M Sharkey (2006), A decade of research and development in disability, virtual reality and associated technologies: promise or practice? *Proc. 6th Intl. Conf. Disability, Virtual Reality & Assoc. Tech.,* Esbjerg, Denmark, pp. 3–16.

C Hendrix and W Barfield (1996), The sense of presence within auditory virtual environments, *Presence Teleoper Virtual Environ*, **3**, pp. 290–301.

T C Hudson, A Seeger, H Weber, J Juliano and A T Helser (2001), VRPN: a device-independent, network-transparent VR peripheral system, *Proceedings of the ACM symposium on Virtual reality software and technology*, pp. 55–61, http://www.cs.unc.edu/Research/vrpn/.

O Lahav and D Mioduser (2002); Multisensory virtual environment for supporting blind persons'acquisition of spatial cognitive mapping, orientation, and mobility skills, *Proc. 4th Intl. Conf. Disability, Virtual Reality & Assoc. Tech.,* Veszprem, Hungary, pp. 213–220.

I Milne and G Rowe (2001), OGRE-3D Program Visualization for C++', *Proceedings of the 3rd Annual LTSN-ICS Conference*, http://www.ogre3d.org/.

T D Parsons and A A Rizzo (2008), Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: a meta-analysis, *Journal of Behavior Therapy and Experimental Pyschiatry*, **39**, pp. 250-261.

T O Rentz, M B Powers, J A J Smits, J R Cougle and M J Telch (2003) Active-imagianal exposure: examination of a new behavioural treatment for cynophobia (dog phobia), *Behaviour Reasearch and Therapy*, **41**, 1337–1353.

G. Riva (2005), Virtual Reality in Psychotherapy: Review, *Cyberpsychology & Behavior*, **8**, pp. 220–230.

J H Sanchez (2004), AudioBattleShip: blind learners cognition through sound, *Proc. 4th Intl. Conf. Disability, Virtual Reality & Assoc. Tech.,* Oxford, UK, pp. 199–206.

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

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

C Suied and I Viaud-Delmon (2008), The role of object categories in auditory-visual object recognition, *9th International Multisensory Research Forum*, Hamburg, Germany, http://imrf.mcmaster.ca/IMRF/ocs/index.php/meetings/2008/paper/view/417

R H J Van der Lubbe and A Postma (2005), Interruption from irrelevant auditory and visual onsets even when attention is in a focused state, *Exp Brain Res*, **164,** pp. 464–471.

D Vastfjall, P Larsson and M Kleiner (2002), Emotion and auditory virtual environments: affect-based judgments of music reproduced with virtual reverberation times, *Cyberpsychol Behav,* **5**, 19–32.

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

I Viaud-Delmon, O Warusfel, A Seguelas, E Rio and R Jouvent (2006), High Sensitivity to Multisensory Conflicts in Agoraphobia Exhibited by Virtual Reality. *European Psychiatry*, **21**, pp. 501–508.

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