Noname manuscript No.
(will be inserted by the editor)

Assisting therapists in assessing small animal phobias by computer analysis of video-recorded sessions

Vicente Castelló · V. Javier Traver · Berenice Serrano · Raúl Montoliu · Cristina Botella

Received: date / Accepted: date

Abstract Behavioural Avoidance Tests (BATs) are commonly used for assessing phobias. While easy to deploy, these tests have some practical difficulties. For instance, therapists have to make distance estimations that are hard to do with accuracy and objectivity; or information regarding the performance of the patients (e.g. their walking pattern) is lost. To alleviate these difficulties, a computerized tool has been developed to extract the walking pattern of patients while approaching the phobic stimulus. From a video-recorded BAT session, two visual representations have been explored to compactly summarize the patient's behavior: a static one (an image) and a dynamic one (an animation). A proof-of-concept prototype has been tested with 23 therapists. Most of the therapists preferred the animated representation, since it provides them with a better sense of the dynamics of how the patient really behaved. The participants agreed that this tool might be useful in assisting therapist when assessing phobia through BATs, since diagnostics could be made in a more accurate and objective way.

Keywords Small-animal phobia · computer vision · visual representations · therapists · clinical diagnostics

V. Castelló · V. J. Traver · R. Montoliu
Institute of New Imaging Technologies (iNIT), Universitat Jaume I, Castellón, Spain
Tel.: +34 964 728727
Fax: +34 964 728730
E-mail: vicastello@gmail.com, {vtraver,montoliu}@uji.es

B. Serrano · C. Botella
Lab. of Psychology & Technology (LABPSITEC), Universitat Jaume I, Castellón, Spain
Tel.: +34 964 387647
Fax: +34 964 729267
E-mail: {bserrano,botella}@uji.es

1 Introduction

Specific phobias involve a marked and persistent fear of a specific situation or object. These disorders, which include phobias to small animals, are very common among the population: some years ago, it was estimated that their prevalence in the United States was of 12.5% of the population, and it was the third most prevalent disorder after depression and alcohol abuse [12]. Specifically in Europe, it is estimated that 22.7 million people are affected, resulting in high costs for the governments. For example, in 2010 the total cost for the treatment of this problem was €19,312 million [18]. All this reflects the need of developing inexpensive and effective assessment and treatment methods for this problem.

The Behavioural Avoidance Test (BAT) [30] is widely used for the assessment of specific phobias (e.g. to small animals) and other anxiety disorders in an effort to provide a common and easy procedure for measuring avoidance behaviours. In this test, the patient is asked to enter a room, and approach the phobic stimulus as much as possible. Her/his performance is scored following an established protocol [30,3]. This test is regularly used as a pre-test for assessing the avoidance and anxiety levels of the patient regarding the phobic stimulus, and as a post-test for assessing the effects of a treatment. Although technically simple, this test suffers from several disadvantages in practice, mainly due to their lack of standardization [16]. On the one hand, some items in the test, such as the self-reported patient's anxiety and the therapist's own estimation, are rated subjectively. On the other hand, other items, such as the proximity of the patient to the target phobic stimulus, can be more objective yet difficult for the therapist to estimate or measure. Fortunately, some of these measurements (mainly the proximity) can be automated, thus bringing objectivity and standardization to the test. One particularly useful technology for this purpose is computer vision. In contrast to some other sensory modalities, the biggest advantage of visual sensors (cameras) is probably that they are unobtrusive for both the therapists and patients. Another benefit is that computer vision-based solutions can be quite inexpensive, especially if camera equipment is already available, as it is often the case in many psychology labs and clinics.

Very little work has been developed regarding the design of computer technologies for supporting the assessment of patients' phobias using BAT. One proposed computer technology consists of using slides and videos of phobic stimuli in a computer-based BAT with the purpose of replacing the use of in vivo stimuli [16]. Other works have used virtual reality BAT for assessing spider phobia [17], one benefit being the experimental replicability. Other reported advantages of the use of technologies are decreases in assessment time and personnel cost, as well as better control and ecological validity [25,23].

It is worth noticing that most of the works have focused on the technological support for the treatment procedure itself. In this regard, web-based applications [4] and virtual and augmented reality systems [19,5,2,28], have been developed and tested and, more recently, serious games on mobiles phones have been explored [3]. The authors of these papers have shown that computer-mediated therapies are effective alternatives to traditional in vivo exposure, while usually providing privacy and bringing higher acceptance and engagement of the patient in the therapy. Although these technologies may involve higher development costs and training effort, they also provide higher effectiveness during assessment and treatment. Therefore, once the system is complete, this kind of tool can be used many times, thus becoming very cost-effective; therapists would be assisted during the test itself and/or afterwards, and could explore visually the behavioral performance or its temporal change.

Regarding computer vision systems applied to human sensing, some work has been done in the past. For instance, facial analysis has been performed in affective computing [27,1,15] and beauty assessment [10]. A very important scope of visual monitoring of people is that of health care [13], which is also the general concern of our work. Recent effort has been paid to early detection of autism by full-body analysis [22,11], and analyzing skeleton kinematic data from movement of children affected by Hemiplegia [21]. Parkinson and other diseases can be recognized from gait patterns [20]. By tracking the body movements, children with cognitive or physical impairments can naturally interact with a computer system [14]. Non-verbal cues, including visual ones, can be used for partly automating the analysis of social behaviours in small groups such as meetings [9]. Therefore, the growing maturity of computer-vision research, the non-invasive nature of these sensors, as well as the flavour of recent work suggest the suitability and relevance of the computer vision technology, and promise an increasing trend in their applicability to psychology.

In this paper, a computerized tool has been developed to extract the walking pattern of patients while approaching the phobic stimulus. Then, two visual representations have been explored to compactly summarize the patient's behavior. In contrast to previous works which regard the *patient* as the primary user of the system, our work is novel in that it uses computer vision to assist the *therapists* in phobia assessment using the BAT. This kind of tool can simplify the work of therapists by providing objective measures of the distance that the patient has walked towards the target, at which speed, and the time spent in getting to it. As discussed above, this represents a step in improving the standardization of phobia assessment using BATs. To the best of our knowledge, this is the first time computer vision is applied in the field of small-animal phobia for this particular problem and these target users (therapists rather than patients). A proof-of-concept prototype has been developed and evaluated in a user study with 23 therapists. The purpose of the prototype was two-fold: to evaluate its usability, utility, and acceptance for psychological assessment, and to find new insights to inform the design of a subsequent, improved system.

Fig. 1 A frame of a video of a simulated (non-real patient) session in the same room where the actual BAT sessions were recorded. The box with the green cover on the table (on the left) represents the target and contains the live animals.

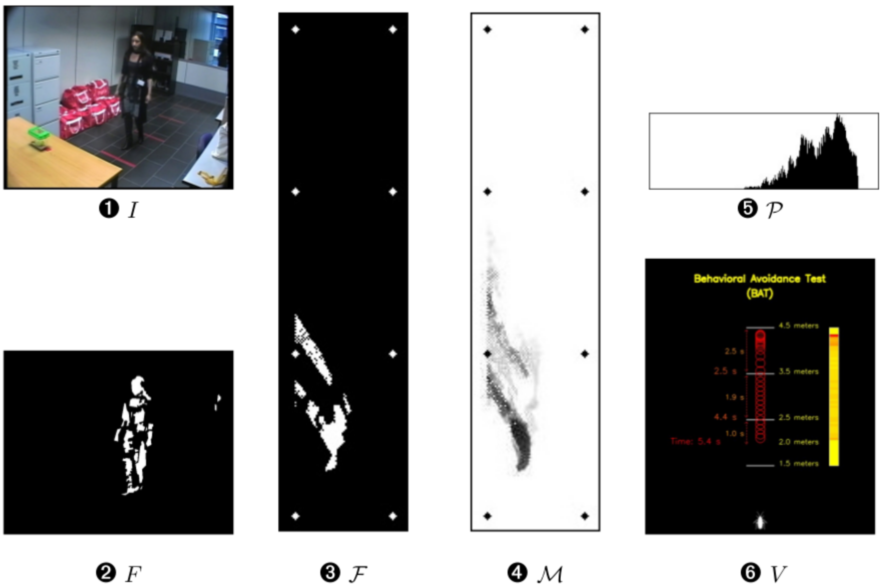


Fig. 2 The pipeline of the system, with steps numbers. See text for some details

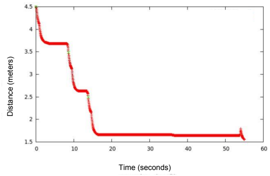
2 System description

Along this section we will use the terms *subject*, *target*, and *distance* to refer to the patient, the phobic stimulus, and the distance he/she walks, respectively.

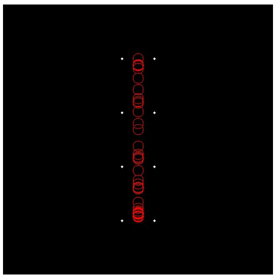
The main purpose of the developed system is to estimate how close to the target phobic stimulus the subject gets or, equivalently, the distance walked from an initial position. A rough camera position-orientation is assumed so that, as a subject walks towards the target, her/his feet can be used as an estimate of the body position (an illustrative frame of a video sequence is given in Fig. 1). To facilitate the automation of the visual processes at this prototype stage, some red marks were placed on the floor in the room. Briefly, the pipeline of the system, detailed in a report [6], and overviewed in Fig. 2, is as follows. From the input frame I, the foreground F is detected with background subtraction [29], which is then rectified $\mathcal{F}$ to the ground plane via homography estimation. Inspired by the idea of motion history [7], a 2D occupancy map $\mathcal{M}$ is built to keep track of recent positions of the subject on the ground plane. Next, this 2D map is projected to a 1D occupancy profile $\mathcal{P}$, which is further processed and filtered to estimate the distance to the target at every frame.

The above procedure provides the actual distance d walked from the starting position, and the remaining distance d' to the target. The time $t(d)$, in seconds, taken to walk a distance d, is computed from the frame number f at which d is estimated, and the video frame rate r (frames per second), as $t(d) = f/r$. The estimated distance d' and the time $t(d)$ spent are useful measures *per se*, yet very crude information for the therapists' usage. Therefore, we looked into ways to provide some richer representation that summarize the subject's behaviour and can be more helpful for the therapists. For instance, whether the subject stopped at some point can be insightful for a diagnosis, but this information is lost if only the pair (d, t) is reported.

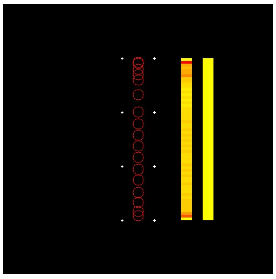
Following the philosophy of user-centered design [24,8], general usability principles were considered, and potential users were regularly consulted, mostly the third author of this paper, who is a therapist herself with experience in conducting BAT sessions. Hence, human-readable friendly representations were explored and, after other considerations, we decided to display the subject at some time intervals with a circle roughly representing a top view of a person. The denser the circles are shown at a point, the slower the subject was moving at that point (circles at different densities can be seen at the centre part in Fig. 4). We chose to draw one circle every five frames, and to use a 5-cm resolution for the space-time scale so that a good clarity-continuity trade-off was achieved. To improve the smoothness of these representations, a Gaussian filtering was first applied to the $d(t)$ signal. A drawback of using circle concentrations is that different stopping/moving times look similar to each other and are therefore difficult to distinguish. To fix this difficulty, a separate colour-coded space-time scale was added (it is the bar on the right-hand side in Fig. 4). The values in this space-time scale for a given video were made relative to the total time spent by the subject in that video. As a further help, complementary textual time information was added next to the space-time scale, at carefully selected time positions. After several iterations and design decisions (a summary of which is given in Fig. 3), the final version is the one shown in Fig. 4.



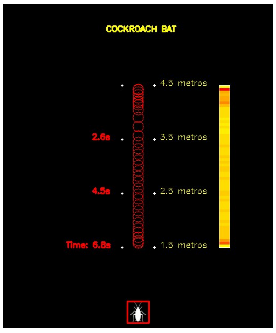
- ➜ Simple distance-time plot
- 👍 Technically accurate and complete
- 👎 Interpretation of behavior not straightforward
- ➜ It was soon discarded



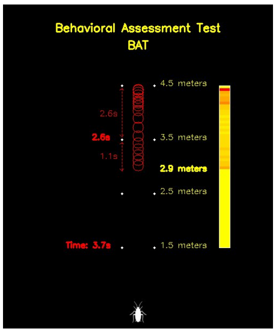
- ➜ Circles as top-view of patients while walking
- ➜ The slower the patient's speed, the more circles
- 👍 Much more intuitive and easier to interpret
- 👎 Different stopping times produce similar circle concentrations
- ➜ Additionally color coding circles by time spent was tried, but was not convincing either



- ➜ Independent color scale to represent time
- ➜ Tradeoff choice decided for space-time granularity for circles and scale
- ➜ Data smoothed for continuity and aesthetic purposes
- 👍 Relative time scale finally preferred over absolute one
- 👎 Relevant absolute times missing



- ➜ Text for times and distances added
- ➜ Different background colors and aesthetic details tested
- 👍 Information of times available
- 👎 Times from one mark to the next were missed



- ➜ Enriched textual information
- 👍 Times between marks available
- 👍 Other details in final prototype (Fig. 4)
- 👍 Dynamic representation considered to overcome potential limitations of the static one

Fig. 3 Some intermediate prototypes produced through iterative design and feedback from the therapists. (This figure is better seen in color)

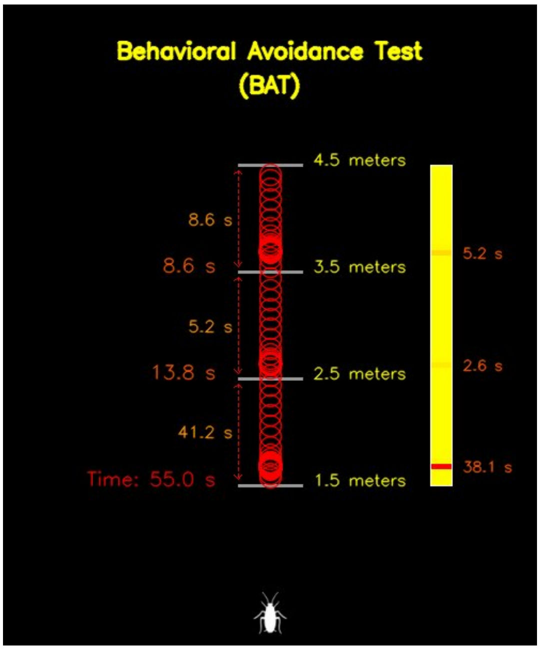


Fig. 4 Final design for the visual summarization automatically built from an input video. Red circles represent the progress of the subject towards the target. The target is represented by the animal drawing in white at the bottom. The yellow-red bar at the right is the space-time scale. Textual annotations are provided for the partial and final time and distance values, as well as for the time peaks. (This figure is better seen in color)

While this "static" visual representation conveys all the information regarding times and distances, we thought that an animation where the elements appeared dynamically as time passes could be useful. We hypothesized that despite watching this animation may require more time than watching the static representation, this effort may pay off since a better understanding of the events might be possible due to higher awareness of the dynamics and the passing of time. Since it is not clear whether the static or the animated visual representation is better than the other (both have advantages and drawbacks), we investigated which one is most useful or preferred by the therapists. These visual representations will be referred to as V_S or V_A, for the static and animated (or dynamic) representations, respectively.

3 Experiments design

We were interested in getting feedback from therapists about the potential benefits and limitations of the visual representations generated from the video-recorded BAT sessions, as well as finding out whether the static or the dynamic representation can generally be preferable.

The participants of the study were $N = 23$ therapists (7 males), all of them psychologists, aged 20–50, with different experience levels (from graduate to doctorate degrees). Since one of our goals was to explore whether the static or dynamic representation can be more helpful than the other one, a between-subjects experiment design was chosen, and participants were randomly assigned to the two experimental conditions considered, either the static (11 therapists) or the dynamic one (12 therapists). First, a warm-up BAT test was used where the participants watched the same video-recorded session and filled in a conventional paper-based BAT questionnaire [30]. Then, seven summarizations, each corresponding to a real case, were tested. To avoid order effects, these cases were presented in random order to each participant.

After the warm-up test, the session consisted of three parts: an estimation task, an opinion survey, and a final interview. The aim of the estimation task was to compare the assessment made by the therapists with the traditional BAT with the assessment made by the therapists using the visual representation for determining to what extent both assessment methods agreed.

The therapists were asked to evaluate, from the visual representation, the quality of the patient's execution for each of the seven cases. To do this, the therapists were asked the following questions: (1) to which degree the patient had approached or whether he/she interacted with the phobic object (4-level *proximity* estimate), and (2) the patient's fear intensity of the subject (3 levels of fear *severity*). To ensure the therapists would understand the visual representation as well as the questions and possible answers, a preview example (we call it "Case 0") was shown prior to the actual seven cases. They were also asked how confident (either low, medium, or high) they were in each of their estimations. We were well aware that this estimation task was difficult to perform accurately since, as stated above, the current visual representation includes only up to the moment the patients get to the last visible floor mark. Notwithstanding this difficulty, we considered this user task useful to place therapists in a realistic situation so that they could better judge the usefulness (or lack thereof) of the visual representation in practical settings.

Therefore, only *after* completing these estimations, an opinion questionnaire was given with 10 items, each with a 5-point Likert scale. The questions tried to capture the participant's experience using the visual representations, their perceived usefulness and their intuitions on their potential benefits and challenges.

After these questions, each participant was shown an example of the visual representation they had not used during the estimation task, and that they actually even ignored it existed. Thus, participants under V_A were shown a V_S representation, and vice versa. Then, they were asked which one they preferred. Finally, an open-ended interview was conducted to further elicit their opinions (both positive and negative) and ideas for the improvement of the visual representation, as well as for envisioning future use scenarios that could inform the subsequent system design.

Table 1 Statistics on estimations. The number of right answers for proximity (p) and severity (s) estimations for each case in relation to the actual measures are given. The answers are considered right according to the actual reference measurement made with the traditional BAT.

		V_A ($n = 12$)		V_S ($n = 11$)		Total ($N = 23$)	
Case	What does the patient do?	p	s	p	s	p	s
1	Walks without stopping	5	7	3	6	8	13
2	Walks without stopping	3	1	1	0	4	1
3	Performs several short time stops while walking	1	2	3	5	4	7
4	Stops for a long time 3 times before getting to the last floor mark	0	0	1	0	1	0
5	Stops right before the second floor mark	12	12	11	11	23	23
6	Walks without stopping	3	9	6	6	9	15
7	Walks quickly without stopping and shows no sign of hesitation	8	9	8	5	16	14

4 Results and discussion

4.1 Estimation task

Although, as discussed above, the estimation task was not a goal in itself, we analyzed the results nonetheless since they can provide some hints on the potential value of the visual representations or the underlying computer data. The results on this task are summarized in Table 1. Please, notice that although these cases correspond to different patients, the description in same cases (1, 2 and 6) is the same because their behaviour was very similar. Although not surprisingly the number of correct estimations is not too high, there are a few interesting issues. In Case 5, all therapists, regardless the visual condition they were assigned to, made the right estimation. This case was very easy because the subject leaves the room before crossing the second floor mark, and this behaviour could be clearly inferred from both visual representations. Case 7 was another easy one, where the subject goes forward quickly without showing signs of fear; about 70% of the therapists guessed correctly. In Case 4, however, all but one therapists made a “wrong” estimation. The reason for this is that therapists understandably considered the frequent and long subject’s stops as a sign of fear, whereas the subject eventually was able to get to touch the plastic container. Another aspect worth noticing is that the kind of visual representation has no clear effect on the estimation accuracy.

Participants with the static representation completed the estimation task faster than the participants with the animated representation, as found with a Mann-Whitney U test ($U = 30.00$, $z = -2.21$, $p < 0.05$, $r = -0.46$). A per-case analysis reveals that the difference occurs, not surprisingly, in the cases with lengthier animations (the longest one is about 1 minute long). Interestingly, if the duration of the animation is subtracted from the time the therapists devoted to the estimation task, then no statistical difference results. This implies that the time is actually spent in watching the animation and the effort to make a decision is similar in both conditions. Indeed, we observed that some therapists under V_A tended to write down their estimation long before the animation stopped playing.

We wondered whether the amount of experience as a psychologist had some influence on any of the variables considered in the user study (e.g. times to complete the estimation task, perceived usefulness of the visual representation, etc.). No statistical significant difference was found, which may suggest that the developed tool can be found useful regardless of the experience level of the therapists.

4.2 Opinion Survey

Interesting findings emerged from the results of the qualitative survey in relation to the usability of the tool. Therapists stated that the visual representation is intuitive and easy to understand (M=4.1, SD=0.6). They also highlighted its foreseeable usefulness in the clinic scope (M=4.0, SD=0.7), the research scope (M=4.7, SD=0.5), and training scope (M=4.2, SD=0.9).

Although both visual representations require some time to be analysed, therapists pointed out that this time is worth spending, compared to only watching original video-recorded BAT session (M=4.1, SD=0.8) or to only using in vivo BAT (M=4.0, SD=0.7). Moreover, they found that this representation provides additional information which is not easily available with current resources (M=4.4, SD=0.8). As a result, therapists agreed that this kind of visual representation would allow them to get a more reliable diagnostic assessment (M=4.1, SD=0.8) and that assessments across different therapists would be more consistent (M=4.3, SD=0.7). Overall, the therapists would really use this kind of visual representation as a support tool in their actual job when using BAT (M=4.5, SD=0.7).

Besides these global statistics, we were interested in finding whether the visual representation made any difference. The only statistical difference resulted in the question on the clinical utility, where therapists under V_A showed higher agreement ($U = 35.50$, $z = -2.05$, $p < 0.05$, $r = -0.42$).

4.3 Final Interview

4.3.1 Preferences

As explained previously, after the opinion survey, participants were shown the alternative representation and asked which one they would prefer. Most therapists (61%, 14/23) chose the animated visual representation as their preferred option, regardless of whether they were assigned V_A or V_S during the estimation task (see Table 2 for details). As nicely put by one participant, motion can

Table 2 Frequency distribution of preferred representation

Assigned condition	Preferred representation		
	V_A	V_S	indifferent
V_A $(n = 12)$	7	2	3
V_S $(n = 11)$	7	2	2
Total $(N = 23)$	14	4	5

just be *intuited* with the static representation whereas it is readily *revealed* by the animated representation. Those therapists who chose V_A explained that they found the animation more useful than the static representation because the dynamics of the patients' movement patterns (e.g. duration of stops or walking speed) allows them to get a better gist of the patients' behaviour than if just a static summary, however complete, is analysed. Only 17% (4/23) of the therapists chose the static representation because they think it is a cost-effective way of analysing all relevant information at a glance. The remaining ones (22%, 5/23) did not find that one representation was particularly more convincing than the other. These findings not only confirm our hypothesis about the potential usefulness of the dynamic representation, but also they are in agreement with the difference reported above on the clinical utility: more people in the V_A condition tended to consider this representation helpful for more accurate diagnosis.

Therapists were asked whether and how they would use these visual representation in real scenarios of clinical practice. Up to 95% of them showed their willingness to use it since they believe the visual representation provides relevant task information in a compact and visually meaningful way, providing accurate and useful information about times and distances. About half of them would prefer to use the visual representation first and the recorded video afterwards, and the other half preferred the other way around, depending on what kind of information they would be seeking. Participant 22 (under V_S) provided one of the most interesting explanations for watching first the visual representation: it would give the concrete details *without* revealing unnecessary personal information on the subject such as gender or age; later on the original video would help him complement information with subtleties such as gestures to better infer nervousness, for instance. Thus, this participant seems to want to avoid being biased in his judgement by looking at the (anonymized) representation first. This observation relates to diminish the subjectivity of BAT tests, which is one of the main purposes of this work.

4.3.2 Benefits and limitations

Ideas and comments therapists mentioned during the interview or that we observed during any part of the study, are summarized in Table 3, where ideas are sorted by their frequency. The benefits perceived point to the system acceptance and its utility in real clinical scenarios. Overall, therapists find that the tool would allow them to perform their work more objectively and productively. Regarding the limitations, therapists tended to miss features which were not planned at this development stage, in particular, displaying patient's interactions with the phobic stimulus. They also raised some representation or visual design issues. For instance, as observed during the preview case (Case 0), many therapists got confused on the colour-coded space-time scale. A possible reason for this confusion is that, in this particular case, the scale had no time annotation (since the subject did not stop for long at any moment). However, when participants were told about the true meaning of this scale they quickly understood it and had no problem in subsequent cases.

A few therapists made some minor design suggestions regarding the use of alternative colours (e.g. for the circle or the background), or replacing the circles with a more human-like appearance of the subject. They generally wished visual representations of the new features they mentioned (e.g., emotional states might be displayed with simple, symbolic faces).

Finally, we selected a few excerpts from the interviews to provide qualitative perception that complement or emphasize previous ideas. Some ideas pointed to current limitations or possible improvements,

> *"It always shows a rectilinear path, not the actual path of the patient"* (Participant 1, V_A)

> *"The representation lacks information that the original video does have, such as emotional responses or face expressions"* (Participant 16, V_S)

> *"It's not clear how to predict the level of anxiety from the information that the representation is conveying"* (Participant 16, V_S)

and some others were very encouraging:

> *"This tool lends itself to be used as part of an on-line phobia diagnosis system"* (Participant 19, V_A)

> *"I find the tool extremely useful. It gives you confidence that you are not missing any relevant detail"* (Participant 23, V_A)

> *"I'd love to have and use this tool. It's not just another tool, it's an essential one"* (Participant 23, V_A)

5 Scope, limitations, and future work

As a research prototype, the developed system revealed to have some limitations, but also significant potential. They are discussed in the following paragraphs.

The red bars on the floor allowed us to simplify the prototype development from the computer vision perspective without this preventing us from exploring the relevant issues of our study. Similarly, this simplification does not represent a practical insurmountable obstacle either, since it can reasonably be overcome with more sophisticated techniques. Interestingly, as a side effect of the presence of these bars, we observed that some patients tended to use them as intermediate milestones towards the greater goal of getting next to the phobic stimulus. This observation might have some implications on the therapy procedures that might be explored as future work.

The use of the now popular RGB+D sensors has not yet been explored because we were primarily interested in using the existing video recordings and camera setup, and we did not need further data either. Certainly, the availability of reliable depth information and even articulated skeleton data might prove helpful for extracting richer behavioral data which might be of clinical relevance.

The expected real benefit of the tool. Although therapists claimed to find the idea of the prototype useful, it might not be completely clear why it is the case. In comparison with employing a traditional BAT, the proposed tool can clearly provide higher objectivity, reliability, and consistency: since the distance walked by the patient is estimated automatically, the therapists are freed from performing this estimation themselves, which brings effectiveness (less work to be done, and getting rich output with minimum effort); reliability and objectivity (the tool offers higher accuracy than the very coarse estimation therapists might do); and consistency (with the help of the tool, different therapists would end up with a more similar assessment of the same or similar clinical cases; and the same therapist would also be more consistent over time). Another clinical benefit is the possibility for the patient to "visualize" his/her own behavior before and after the treatment. For example, the patients could see (a representation of) themselves exhibiting a safer approach to the phobic object, which can reinforce the therapeutic value of the treatment. If the tool is used (long) after the session, by the same or different therapists, the fact of having a simple graphical representation of the patient (such as circles representing a top view) would bring objectivity, since personal traits of the patients (gender, age, emotional state) are concealed. As a result, therapists can focus on the relevant variables only and avoid being biased by other irrelevant factors.

Usage scenarios of this tool might be envisioned in the clinic, training and research scopes, as suggested previously in the paper. The benefits in the clinic scope has already discussed above. Yet another possibility would be to provide a greater automation of the comparison between the pre- and post-treatment BAT sessions, thus providing a significant help and time saving as a decision support for the therapist. An example of the interest of this kind of tool for training would be to easily show to Psychology trainees BAT examples corresponding to patients with different degrees of fear, and then discuss a series of individualized exposures for each patient.

Finally, for research purposes, the tool might be useful to explore how the measured variables correlate with real assessment and estimations, so that machine learning can be used for developing a model which is useful for semi-automatic estimation, as part of a decision support system. Many interesting questions can be studied with the help of this tool such as whether the actual video recording help or hinder to yield a more accurate diagnostic in comparison with the visual representation generated with the proposed system.

It can also be envisioned a low-cost easy-to-install commercial kit to support either therapists or lecturers in their respective assessment and training activities.

Adding interactivity While the dynamic representation has been generally preferred, animations may violate the second principle of good graphics, the *apprehension* principle [26]. However, adding higher interactivity to the proposed tool may easily overcome this limitation and provide richer analysis possibilities to the therapists.

Generalization With some changes, a similar tool may be used in the assessment tasks of other psychological disorders. In cases of social anxiety, a visual representation of the patients' peformance and related data may be useful. In schizophrenia, it can help identify interaction patterns and monitor the evolution with the therapy. A third interesting possibility might be in some psychological treatment of children, since the proposed techniques lend themselves to visually entertaining representations that bring some charm to children.

6 Conclusions

Behavioural Avoidance Tests (BATs) are widely known and used for assessing phobias. However, they suffer from some practical difficulties such as the subjectivity of some of its items. This issue is addressed in this work through a computer-vision system that estimates the distances walked by the patients, from the video-recorded BAT sessions. One static and one dynamic visual representations, which summarize the patient's behaviour during the session, have been proposed as a proof-of-concept prototype which has been tested through a user study with 23 therapists.

The feedback provided by the therapists has been positive and encouraging, as well as informative for refinement. They found the visual representations intuitive, and felt that a tool based on these representations would allow them to perform the phobia diagnostics more accurately, and with higher objectivity and productivity. The animated representation has generally been preferred over the static one, since the perception of the dynamics of the walking behaviour becomes much more direct and straightforward. The participants in the study envisioned that this kind of technology would help psychologists in the clinical, training, and research scopes.

Acknowledgements. This work has been partly supported by Fundació Caixa-Castelló (through grant P1-1A2010-11) and Generalitat Valenciana (through grant PROMETEOII2014062).

Table 3 Some (categorized) comments users did during the interview. The number of users (out of $N = 23$) who made each comment is given

(a) Comment on perceived or potential benefits	No. users
It provides precise information on distance and time that can be helpful in BAT assessment	16
Draws attention on relevant or useful movement patterns that would otherwise go unnoticed	13
Lends itself to objective assessment and unified criteria	12
Allows quicker and easier comparisons between pre-test and post-test of the same patient	6
Saves time and effort with respect to viewing the video-recorded session	5
Facilitates the detection of similar behaviours and their relationship with similar avoidance levels	5
The representation is visually clear, simple and easily understandable	4
Would ease tracking the patient's evolution by reminding at a glance sessions long past	2
Could be used as evidence to support the feedback given to patients when following their case	1
Hides real identity of patient leading to unbiased assessment and higher confidentiality and privacy	1
(b) Comment on perceived limitations or drawbacks	No. users
The colour-coded time scale is misunderstood as an anxiety rank	19
The actions of the patient after the last visible floor mark are missed	14
The audio signal or textual descriptions of patient's speech would be beneficial	6
Visual coding of anxiety level or other emotional states would be appreciated	4
Steps backwards or how the patient leaves the room are not captured but would be useful hints	4
Circles are mistaken as actual gait steps rather than time-regular positions	4
Non-verbal cues such as gestures and postures would be informative	4

References

1. Jeremy N. Bailenson, Emmanuel D. Pontikakis, Iris B. Mauss, James J. Gross, Maria E. Jabon, Cendri A. C. Hutcherson, Clifford Nass, and Oliver John. Real-time classification of evoked emotions using facial feature tracking and physiological responses. *International Journal of Human-Computer Studies*, 66(5):303–317, 2008.
2. C. Botella, J. Bretón-López, S. Quero, R. M. Baños, and A. García-Palacios. Treating cockroach phobia with augmented reality. *Behavior Therapy*, 41(3):401–413, September 2010.
3. Cristina Botella, Juani Bretón-López, Soledad Quero, Rosa María Baños, Azucena García-Palacios, Irene Zaragozá, and Mariano Alcañiz Raya. Treating cockroach phobia using a serious game on a mobile phone and augmented reality exposure: A single case study. *Computers in Human Behavior*, 27(1):217–227, 2011.
4. Cristina Botella, María José Gallego, Azucena García-Palacios, Veronica Guillen, Rosa María Baños, Soledad Quero, and Mariano Alcañiz Raya. An internet-based self-help treatment for fear of public speaking: A controlled trial. *Cyberpsychology, Behavior, and Social Networking*, 13(4):407–421, 2010.
5. Juani Bretón-López, Soledad Quero, Cristina Botella, Azucena García-Palacios, Rosa María Baños, and Mariano Alcañiz Raya. An augmented reality system validation for the treatment of cockroach phobia. *Cyberpsychology, Behavior, and Social Networking*, 13(6):705–710, 2010.
6. Vicente Castelló, V. Javier Traver, Berenice Serrano, Raul Montoliu, and Cristina Botella. Computer vision analysis of video-recorded sessions for BAT-based phobia as-

sessment. Technical report, Institute of New Imaging Technologies, Universitat Jaume I, Castellón, Spain, 2015.
7. J.W. Davis and A.F. Bobick. The representation and recognition of human movement using temporal templates. In *Computer Vision and Pattern Recognition*, pages 928–934, 1997.
8. Alan Dix, Janet E. Finlay, Gregory D. Abowd, and Russell Beale. *Human-Computer Interaction (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2003.
9. Daniel Gatica-Pérez. Automatic nonverbal analysis of social interaction in small groups: A review. *Image & Vision Computing*, 27(12):1775–1787, 2009.
10. Hatice Gunes and Massimo Piccardi. Assessing facial beauty through proportion analysis by image processing and supervised learning. *International Journal of Human-Computer Studies*, 64(12):1184–1199, 2006.
11. Jordan Hashemi, Thiago Vallin Spina, Mariano Tepper, Amy Esler, Vassilios Morellas, Nikolaos Papanikolopoulos, and Guillermo Sapiro. A computer vision approach for the assessment of autism-related behavioral markers. In *2012 IEEE International Conference on Development and Learning and Epigenetic Robotics, (ICDL-EPIROB 2012)*, pages 1–7, 2012.
12. R. C. Kessler, P. Berglund, O. Demler, R. Jin, K. R. Merikangas, and E. E. Walters. Lifetime prevalence and age-of-onset distributions of DSM-IV disorders in the national comorbidity survey replication. *Archives of General Psychiatry*, 62(6):593–602, 2005.
13. Ben Kröse, Tim Oosterhout, and Tim Kasteren. Activity monitoring systems in health care. In Albert Ali Salah and Theo Gevers, editors, *Computer Analysis of Human Behavior*, pages 325–346. Springer London, 2011.
14. Cristina Manresa Yee, Joan Jordi Muntaner, and Diana Arellano. A motion-based interface to control environmental stimulation for children with severe to profound disabilities. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 7–12, 2013.
15. D. McDuff, R.E. Kaliouby, and R.W. Picard. Crowdsourcing facial responses to online videos. *IEEE Trans. on Affective Computing*, 3(4):456–468, 2012.
16. Charles Teng Tat Meng, Kenneth C. Kirkby, Frances Martin, Lisa J. Gilroy, and Brett A. Daniels. Computer-delivered behavioural avoidance tests for spider phobia. *Behaviour Change*, 21:173–185, 8 2004.
17. A. Mühlberger, M. Sperber, M.J. Wieser, and P. Pauli. A virtual reality behavior avoidance test (VR-BAT) for the assessment of spider phobia. *Journal of CyberTherapy & Rehabilitation*, 1:147–158, 2008.
18. J. Olesen, A. Gustavsson, M. Svensson, H.-U. Wittchen, and B. Jönsson. The economic cost of brain disorders in Europe. *European Journal of Neurology*, 19(1):155–162, January 2012.
19. T. D. Parsons and A. A. Rizzo. Affective outcomes of virtual reality exposure therapy for anxiety and specific phobias: A meta-analysis. *Journal of Behavior Therapy & Experimental Psychiatry*, 39:250–261, 2008.
20. Bogdan Pogorelc, Zoran Bosnić, and Matjaž Gams. Automatic recognition of gait-related health problems in the elderly using machine learning. *Multimedia Tools & Applications*, 58(2):333–354, 2011.
21. Md. Abdur Rahman. Multimedia environment toward analyzing and visualizing live kinematic data for children with hemiplegia. *Multimedia Tools & Applications*, 74(15):5463–5487, 2015.
22. James M. Rehg. Behavior imaging: Using computer vision to study autism. In *Proc. of the IAPR Conference on Machine Vision Applications (IAPR MVA 2011)*, pages 14–21, 2011.
23. G. Riva. Virtual reality as assessment tool in psychology. *Studies in health technology and informatics*, 44:71–79, 1997.
24. Helen Sharp, Yvonne Rogers, and Jenny Preece. *Interaction Design: Beyond Human Computer Interaction*. John Wiley & Sons, 2007.
25. T. J. Trull. Expanding the aperture of psychological assessment: introduction to the special section on innovative clinical assessment technologies and methods. *Psychological Assessment*, 19(1):1–3, 2007.
26. Barbara Tversky, Julie Bauer Morrison, and Mireille Betrancourt. Animation: can it facilitate? *International Journal of Human-Computer Studies*, 57(4):247–262, 2002.

27. Michel F. Valstar, Hatice Gunes, and Maja Pantic. How to distinguish posed from spontaneous smiles using geometric features. In *Proceedings of the 9th international conference on Multimodal interfaces*, ICMI '07, pages 38–45, New York, NY, USA, 2007. ACM.
28. Maja Wrzesien, Mariano Alcañiz, Cristina Botella, Jean-Marie Burkhardt, Juana Bretón-López, Mario Ortega, and Daniel Beneito Brotons. The therapeutic lamp: Treating small-animal phobias. *IEEE Computer Graphics and Applications*, 33(1):80–86, 2013.
29. Z. Zivkovic. Improved adaptive gausian mixture model for background subtraction. In *International Conference Pattern Recognition*, pages 28–31, 2004.
30. L.-G. Öst, P.M. Salkovskis, and K. Hellstrom. One-session therapist-directed exposure vs. self-exposure in the treatment of spider phobia. *Behavior Therapy*, 22(3):407–422, 1991.