Judging crowds' size by ear and by eye in virtual reality

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

The final publication is available at Springer via http://dx.doi.org/10.1007/s12193-016-0228-5

Judging crowds' size by ear and by eye in virtual reality

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

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

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

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

*Corresponding author.

Marine Taffou
1 place Igor Stravinsky
75004 Paris – France

mail: marine.taffou@gmail.com

Acknowledgments

This research was supported by the EU FP7-ICT-2011-7 project VERVE (http://www.verveconsortium.eu/), grant n°288910. This work was performed within the Labex SMART (ANR-11-LABX-65) supported by French state funds managed by the Agence Nationale de la Recherche within the Investissements d'Avenir programme under reference ANR-11-IDEX-0004-02. The research leading to these results has also received funding from the program "Investissements d'avenir" ANR-10-IAIHU-06. We thank Thibaut Carpentier and Kévin Perros for their work on the elaboration of the auditory component of the virtual environment. We thank Camille Frey and Cassandra Visconti who contributed to the experimentation.

Conflict of Interest: The authors declare that they have no conflict of interest.

Abstract

Judging the size of a group of people is an everyday task, on which many decisions are based. In the present study, we investigated whether judgment of size of different groups of people depended on whether they were presented through the auditory channel, through the visual channel, or through both auditory and visual channels. Groups of humanoids of different sizes (from 8 to 128) were presented within a virtual environment to healthy participants. They had to judge whether there was a lot of people in each group and rate their discomfort in relation to the stimuli with Subjective Units of Distress.

Our groups of 96 and 128 virtual humans were judged as crowds regardless of their sensory presentation. The sensory presentation influenced participants' judgment of virtual human group size ranging from 8 to 48. Moreover, while the quantity judgments in the auditory condition increased linearly with the group size, participants judged the quantity of people in a logarithmic manner in the two other sensory conditions. These results suggest that quantity judgment based on auditory information in a realistic context may often involve implicit arithmetic.

Even though our participants were not phobic of crowds, our findings are of interest for the field of virtual reality-based therapy for diverse disorders because they indicate that quantity judgment can potentially be altered in a sensory-specific manner in patients with fear of crowds.

Keywords: quantity judgment, sensory modality, auditory-visual, crowd, virtual reality, humanoids

1. Introduction

In the past twenty years, medical application of virtual reality technology to psychotherapy and rehabilitation has addressed a variety of pathologies such as phobias, eating disorders and neurological disorders. Fear of crowds is a symptom found in several of these disorders. Specifically, fear of crowd is observed in anxiety disorders like agoraphobia [1] and is a factor associated with fear of falling in Parkinson disease, multiple sclerosis, and ageing, resulting in a vicious cycle of falls, fear of falling, functional decline, and more falls [e.g. 2]. This symptom is an important contributor to the risk of social exclusion for these populations. New technologies can be developed to support the treatment of people who are at risk of social exclusion because of a fear linked to the number of individuals that can be met outside their home. In this framework, we explored the advantages of auditory-visual virtual environments, displaying ambient sounds and human speech sounds along with the visual stimulation, for the treatment of fear of crowds. The fear of crowds has indeed both visual and auditory relevant components because the humans composing the crowds generate both visual and auditory stimulation. Several studies have used virtual environments containing virtual humans to evaluate the potential of virtual reality-based therapy for panic disorder with agoraphobia [3-9] and social anxiety with fear of public speaking [10-13]. Specifically, some studies showed a reduction of fear in patients suffering from panic disorder with agoraphobia after exposure sessions in virtual environments depicting a shopping mall, a bus or a subway train [4, 8]. However, whereas an effort is made in these studies in rendering accurate visual stimulation, auditory rendering is often neglected. The auditory stimuli are often absent and if present are not rendered in 3D and are not interactive. Consequently, the auditory aspects of the phobia are often underexploited in the therapy and little is known on whether hearing without seeing a crowd (for example if it is behind the corner of the street), seeing without hearing a crowd (if the crowd is silent) or both hearing and seeing the crowd has an impact on patients' fear. In the attempt of investigating the role of the sensory presentation of crowds of humans in the fear of crowds, we needed to determine the quantity of virtual humans that is necessary for a group to be evaluated as a crowd in our virtual environment, and whether this quantity would differ as a function of the sensory presentation of the stimulation.

Choosing which queue to join at the supermarket, which subway car to enter during rush hour or whether to make a detour to avoid a crowd involves the approximation of the number of individuals in a group. Studies investigating numerical estimation have mostly used non-meaningful material (e.g. dots, tones) and have suggested that the representation of approximate numerosities is generated by innate neural mechanisms [e.g. 14-20]. Whereas this representation of quantity is thought to be independent of the sensory modality of stimuli [e.g. 21], a recent study has reported differences in approximate numerical judgments between visual and auditory stimuli. In a numerical comparison task, Tokita et al. observed a better precision with sequences of tones compared to the precision with sequences of flashes [22] and proposed that quantity judgment involves complex processes, which may be influenced by the sensory presentation of stimuli.

Given that it is not yet clear whether quantity judgment is influenced by sensory modality with meaningful stimuli such as groups of individuals, we conducted an experiment using a simple procedure corresponding to a judgment that people do actually make in real life to examine this question. Participants were asked to judge the size of groups of individuals in three sensory conditions: as they could perceive the group of individuals through the auditory channel only, through the visual channel only and through both the auditory and visual channels. We used a virtual reality setup to display groups of virtual humans, which could convey information via both audition and vision, within an auditory-visual virtual environment. Virtual reality techniques integrate real-time computer graphics, body tracking devices and visual and auditory displays to facilitate the presentation of high quality 3D visual and auditory information. Thus, we exploited the unique advantages of virtual reality to display different sizes of groups of virtual humans embedded in a realistic context as well as to manipulate their sensory presentation.

2. Materials and methods

2.1. Participants

Twelve participants (two women; age = 26.50 ± 4.60) with normal audition and normal or corrected to normal vision voluntarily participated in the experiment. They were recruited among the members of the Institute of Research and Coordination Acoustic/Music and among the students of the University Pierre et Marie Curie in Paris. 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.

2.2. Virtual reality setup

The virtual reality 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 1). 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 non-individual Head Related Transfer Functions (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 23, 24]. With this procedure, the virtual sound source location can be manipulated by rendering accurate auditory cues such as inter-aural intensity and time differences as well as spatially dependant frequency spectra. An 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.

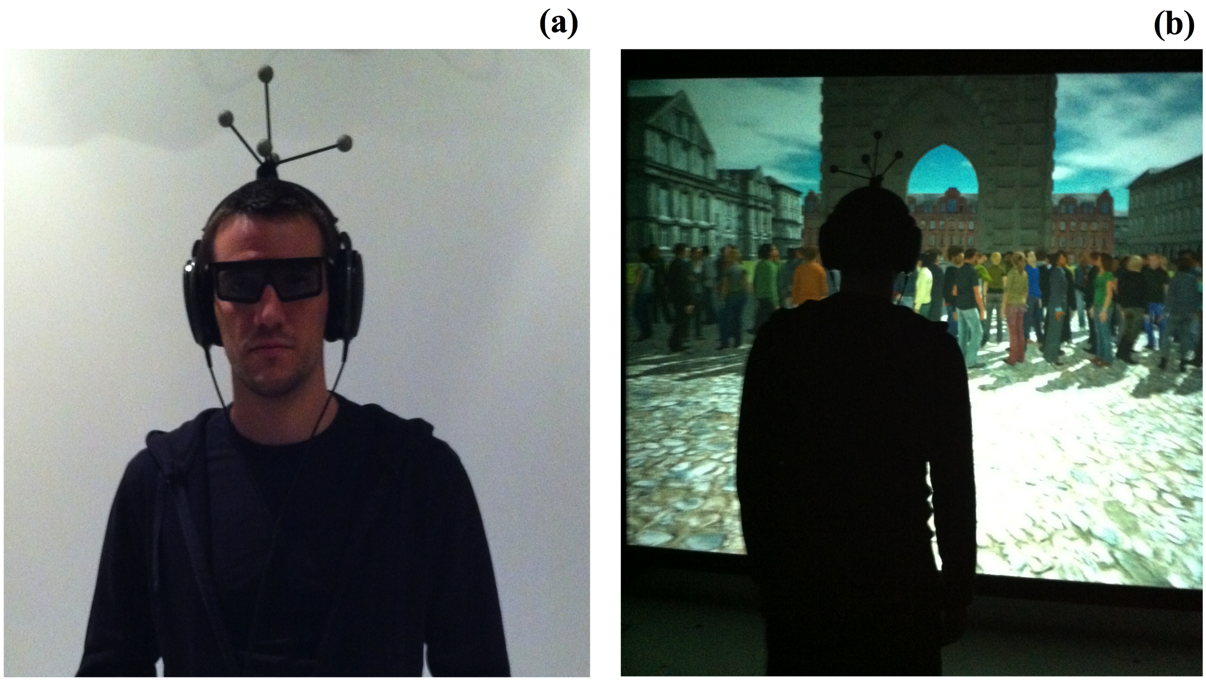

Fig. 1 Virtual reality setup. (a) A participant equipped with polarized glasses, headphones and a tracking device. (b) A participant standing in front of the stereoscopic passive screen during an immersion in virtual reality.

2.3. Virtual environment

The virtual environment (VE) reproduces the outdoor environment of the Trinity College Dublin campus composed of buildings, alleys and vegetation. Animated virtual individuals, referred to as humanoids, can be placed in the VE.

The auditory component of the VE consisted of human speech and of an ambient audio environment composed of bird sounds and of urban activity. The different sound files of human speech were recorded in an anechoic chamber. 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…). Discussions between native French speakers 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 were created by clipping his/her speech. These sound files were then normalized and equalized in terms of loudness and compressed afterwards with standard voice compression parameters.

2.4. Stimuli: groups of humanoids

Seven stimuli were designed (see Figure 2). These seven stimuli consisted of groups composed of different numerosities of humanoids (8, 16, 32, 48, 64, 96 and 128) organized in subgroups of size 1 to 8. All of the humanoids were animated, located at a fixed position and involved in a subgroup discussion as either a talker or as a listener. No humanoid was walking. The humanoids were animated using motions that have been captured from real humans that were talking or listening in a conversation [25]. These motions include head and hand movements as well as postural changes. Talkers were assigned a talking animation and a gender-matching sound file of human speech whereas listeners were assigned only a listening animation. For the talking humanoids, there was no lip synchronization. In order to avoid the technical difficulties of rendering interactive discussions between humanoids and to maintain the characteristics of the simulation stable over time, only one humanoid per subgroup was designated as a talker. Humanoids who were alone (subgroups of one individual), were talkers while speaking on 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 concomitantly with the increase in the numerosity 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.

The position of each humanoid in the scene was defined so that the whole group of humanoids was in the participant's field of view when he/she was standing at 6 meters from the group (which includes the distance of

1.73 from the screen). Groups were composed of an equal number of male and female humanoids 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.

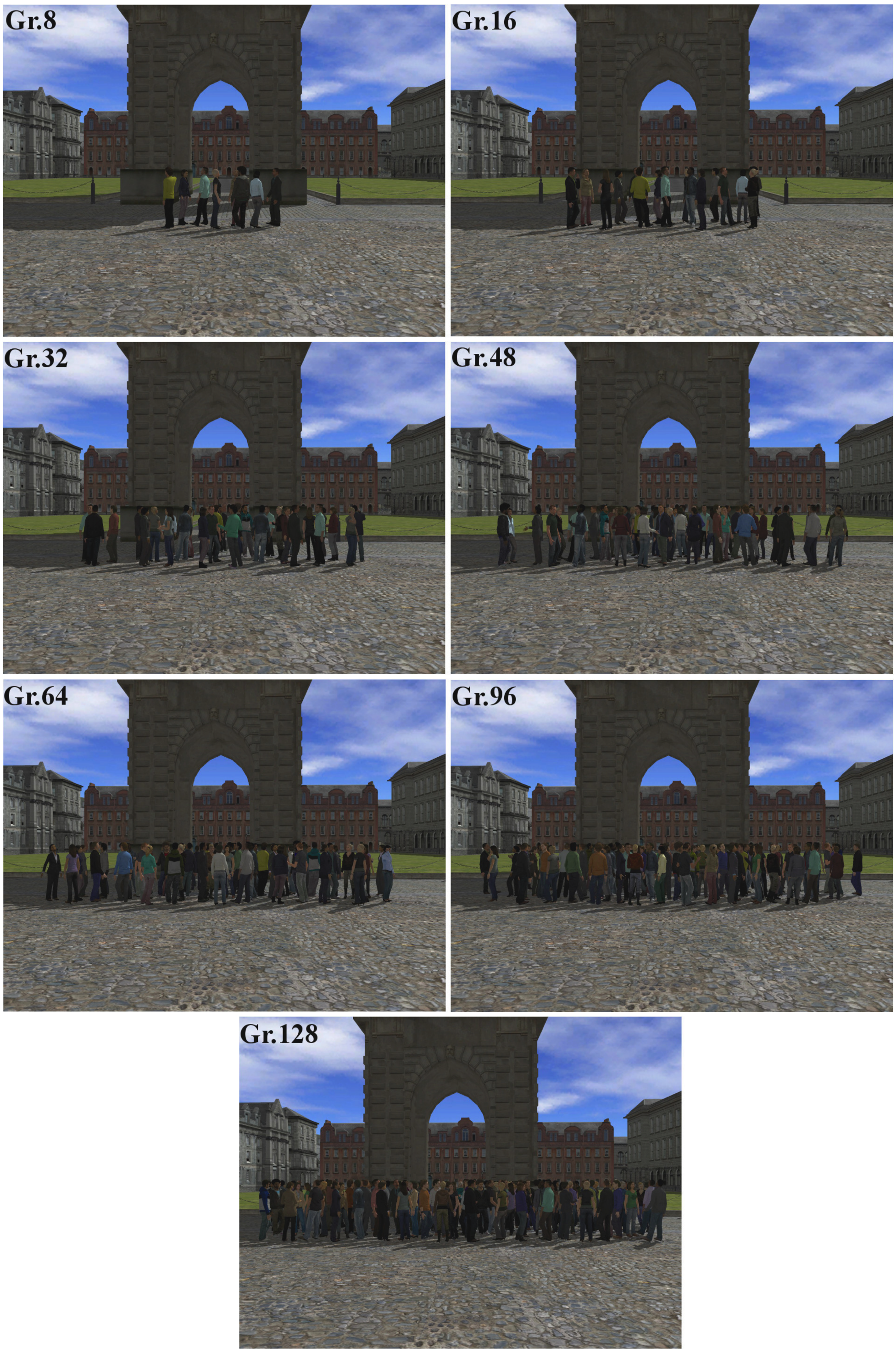

Fig. 2 The seven stimuli used in the study. They consisted of groups composed of different numerosities of humanoids: 8 (Gr.8), 16 (Gr.16), 32 (Gr.32), 48 (Gr.48), 64 (Gr.64), 96 (Gr.96) or 128 (Gr.128) humanoids.

2.5. Procedure

Upon arrival, participants completed the trait portion of the State Trait Anxiety Inventory [26]. They also completed a questionnaire exploring the fear of crowds (https://www.researchgate.net/publication/286780388_Fear_of_crowds_questionnaire), which is a questionnaire that we designed and used to select participants sensitive to the fear of crowd for another experiment. In the current study, this questionnaire was used to verify that none of our 12 participants was sensitive to the fear of crowd. This questionnaire comprises 15 items describing common crowded situations and addressing different aspects of an encounter with a crowd such as the sensory modalities through which the crowd is sensed (auditory, visual, and/or tactile stimulation) or the mobility of the crowd (fixed in space as when "standing in a crowded bus" or in movement as when "walking in a subway station during rush hour"). Each item has to be rated according to the intensity of discomfort they elicit on a scale from 0 (no discomfort) to 3 (extreme discomfort). The minimal total score on this crowd fear questionnaire is 0, with a maximum of 45. Two hundred and twenty-eight individuals (121 women, age: 24.55 ± 5.32) completed this questionnaire. A distribution of scores was obtained and served as a basis to verify that the participants to the current experiment were not sensitive to the fear of crowds. After the completion of the questionnaires, participants completed three immersions in the VE, corresponding to three different sensory conditions. The VE was first presented through the auditory modality only (A), through the visual modality only (V), then through both auditory and visual modalities (AV). During the immersions, participants stood at 1.73m from the center of the screen. The virtual scene placed them in a square of the VE, in front of a big arch (see Figure 1). They did not navigate in the virtual scene. They 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).

Each immersion lasted around 5 minutes and included the presentation of the seven stimuli composed of groups of humanoids, which sizes increased progressively (from 8 to 128: Gr.8, Gr.16, Gr.32, Gr. 48, Gr. 64, Gr.96, Gr. 128). This presentation, with increasing group size, was chosen in analogy with the principle of progressive exposure in therapies for phobias. Given that the final aim of this study is to help designing virtual environments for the treatment of fear of crowd, we used a progressive group size presentation to determine how many humanoids are needed for a group to be considered as a crowd. Each stimulus was presented between the participant and the arch, with the closest humanoids being at 6m (which includes the distance of 1.73 from the screen) from the participant.

For each of the seven stimuli, participants had to indicate how much they agreed 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 quantity estimation of humanoids composing the groups. Participants also indicated the intensity of discomfort they experienced when imagining they had to walk through the arch in the presence of the different stimuli using Subjective Units of Distress (SUD) [27]. SUD is a self-report measurement of experienced discomfort or fear on a 0-10 point scale, which is widely used in behavioural research and therapy [e.g. 28-30].

The order of the three immersions (auditory, visual then auditory-visual) was chosen in order to prevent participants from mentally visualizing the group of humanoids in the A immersion. In the A immersion, 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 immersion 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 and to verify that they actually perceived the sound spatialization and the visual stereoscopy. The duration of the total experiment was about one hour.

3. Results

Participants’ mean score on the fear of crowd questionnaire was 11.33 (SD = 6.62) with a median score of 13.50 (range [1.00; 19.00]). On the basis of the scores obtained by 228 individuals, who completed this fear of crowd questionnaire, we consider that our participants were not sensitive to the fear of crowd. Indeed, none of them had a score superior to the 80th centile of the distribution of the scores in the sample of 228 individuals. Trait anxiety scores ranged from 29 to 51 with a mean score of 38.25 (SD = 9.22) and a median of 39.00. These scores are not extreme and indicate that none of our participants was especially anxious.

3.1. Quantity Judgment

To determine which stimuli can be used as crowd stimuli, we first calculated the mean and the confidence interval 95% of the mean (CI95) of the quantity judgment response 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 quantity judgment response was higher than five. As reported in Table 1, the groups of 96 and 128 humanoids met this criterion in the three sensory conditions.

Table 1 Overview of the results

	Sensory presentation of the group of virtual humans		
	Auditory	**Visual**	**Auditory-Visual**
Group sizes evaluated as crowds	96 and 128	96 and 128	96 and 128
Function shape of judged quantity	Linear	Logarithmic	Logarithmic
Function shape of reported discomfort	Linear	Linear	Linear

We then tested the effect of the sensory condition on the estimation of the quantity of humanoids in response to each of the seven stimuli (Gr.8, Gr.16, Gr.32, Gr. 48, Gr. 64, Gr.96, Gr. 128). An ANOVA was conducted on quantity judgments with the within-subject factors STIMULUS (Gr.8, Gr.16, Gr.32, Gr.48, Gr.64, Gr.96, Gr.128) and SENSORY CONDITION (A, V, AV). The main effect of STIMULUS was significant [$F_{(6, 66)}$ = 101.72, p < .001, η_p^2 = .902]. The two-way interaction STIMULUS* SENSORY CONDITION was also significant [$F_{(12, 132)}$ = 2.59, p = .004, η_p^2 = .190] suggesting that participants' estimation of the quantity of humanoids was differently modulated by the sensory condition as a function of the stimulus. Whereas participants' judgment of the quantity of humanoids did not differ according to the sensory condition for Gr.64, Gr.96 and Gr. 128 (post hoc Newman-Keuls' tests: p > .336 in all cases), the sensory condition had an impact on quantity judgment for the smaller groups of humanoids. The quantity judgment was higher in the auditory condition compared to the visual and auditory-visual conditions for the Gr.8 (post hoc Newman-Keuls' tests: p < .001 in both cases). It was also higher in the auditory than in the auditory-visual condition for the Gr. 16 (post hoc Newman-Keuls' tests: p = .001). For the Gr.32, both the quantity judgments in the auditory and visual conditions were higher than in the auditory-visual condition (post hoc Newman-Keuls' tests: p < .017 in both cases). For the Gr.48, the quantity judgment was also higher in the visual condition in comparison to the auditory-visual condition (post hoc Newman-Keuls' tests: p = .036).

The shape of response functions in experiments on numerical estimation of large numerosities generally is negatively accelerated and fits a logarithmic or power function rather than a linear function [29-31]. We examined whether this property extended to the data sets that we collected in our experiment on quantity judgement with meaningful stimuli in the three different sensory modalities. We tested whether the relationship between the number of humanoids composing the different groups and participants' response on the quantity judgment task was linear or logarithmic. A multiple regression procedure was used to evaluate, separately for each sensory condition, the contribution of a logarithmic regressor [log(group size)], over and above a linear regressor (group size), in the quantity judgment. As shown in Figure 3, there is a significant logarithmic component in participants' responses in the V and AV conditions whereas participants rated linearly the quantity of humanoids in the A condition (see significance levels and regression weights on the figure). These results are consistent with the results of the ANOVA concerning the interaction between sensory presentation of stimuli and their numerosity.

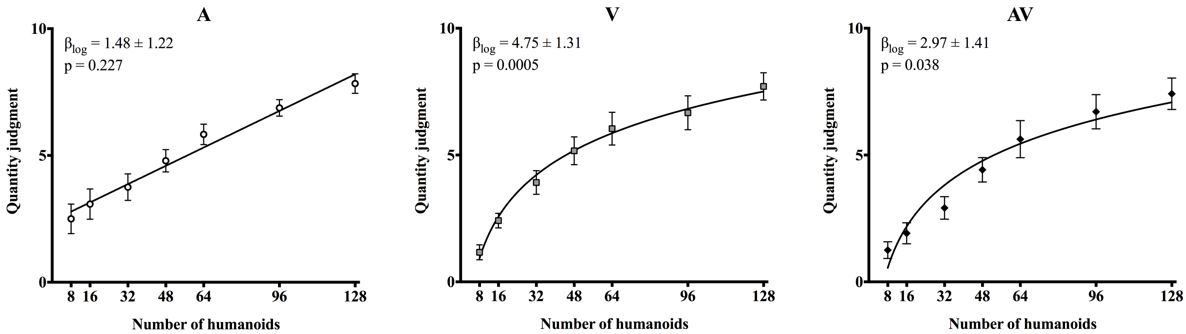

Fig. 3 Participants judged the quantity of humanoids constituting seven different groups (Gr.8, Gr.16, Gr.32, Gr. 48, Gr. 64, Gr.96, Gr. 128) by indicating the level of their agreement with the statement "*There is a lot of people"* on a scale from 0 (*I totally disagree*) to 10 (*I totally agree*) as the groups of humanoids were presented in three different sensory conditions: auditory (A), visual (V) and auditory-visual (AV). This figure reports participants' mean judgement of humanoids quantity (± SEM) in the auditory (left), visual (middle) and auditory-visual (right) sensory conditions as a function of the number of humanoids composing the groups. There was a significant effect of sensory presentation on quantity judgment for the groups with small numerosities (8-48). There was a significant logarithmic component in participants' judgment of the quantity of humanoids in the V and AV sensory conditions whereas they rated linearly the quantity of humanoids in the A sensory condition.

3.2. Experienced discomfort

We tested the effect of the sensory condition on participants' discomfort in response to each of the seven stimuli (Gr.8, Gr.16, Gr.32, Gr. 48, Gr. 64, Gr.96, Gr. 128). Given that the distribution of participants' SUDs deviated from normal distribution, we conducted, separately for each stimulus, non-parametric Friedman ANOVAs on participants' SUDs with the within-subject factor SENSORY MODALITY (A/V/AV). There was no significant effect of SENSORY CONDITION on participants' SUDs in response to each of the different groups of humanoids ($p > 0.157$ in all cases).

We also examined whether the relationship between the number of humanoids composing the different groups and participants' discomfort was linear or rather logarithmic. The multiple regression procedure was used to evaluate, separately for each sensory condition, the contribution of a logarithmic regressor [log(group size)], over and above a linear regressor (group size), in SUDs reported by participants. As shown in Figure 4, the contribution of the logarithmic regressor in participants' responses was not significant in all sensory conditions (see significance levels and regression weights on the figure).

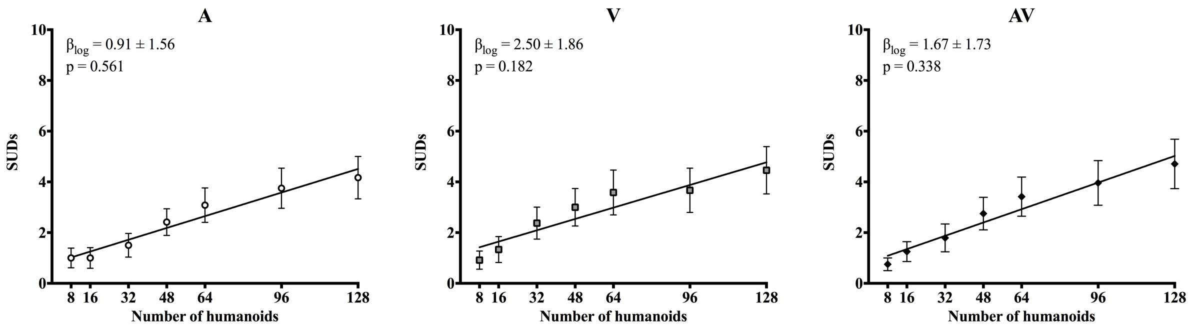

Fig. 4 Participants reported the intensity of their discomfort in response to the seven different groups of humanoids (Gr.8, Gr.16, Gr.32, Gr. 48, Gr. 64, Gr.96, Gr. 128) with Subjective Units of Distress (SUDs) on a scale from 0 (*no discomfort*) to 10 (*high discomfort*) as the groups of humanoids were presented in three different sensory conditions: auditory (A), visual (V) and auditory-visual (AV). This figure reports participants' mean SUDs (± SEM) in the auditory (left), visual (middle) and auditory-visual (right) sensory conditions as a function of the number of humanoids composing the groups. The sensory presentation of the groups of humanoids did not significantly influence participants' experience of discomfort. Participants' discomfort increased linearly rather than logarithmically with the number of humanoids in all sensory conditions.

4. Discussion

We investigated the influence of sensory presentation on the judgment of quantity using meaningful stimuli. We exploited the unique advantages of virtual reality to manipulate the sensory presentation of meaningful stimuli consisting of groups of individuals as well as to display the stimuli embedded in a realistic context. The groups of 96 and 128 humanoids were evaluated as crowds in all three sensory conditions. Participants' quantity judgements were influenced by the sensory presentation only for the stimuli with smaller numerosities (8-48). The relation between quantity judgments and stimuli numerosity was examined with visual, auditory and auditory-visual presentations with a multiple regression procedure to determine whether linear or logarithmic regression provides the best model to describe participants' quantity judgments in each sensory condition. Whereas a logarithmic model better fitted participants' data than a linear model in the visual and auditory-visual sensory conditions, participants rated linearly the magnitude of stimuli sizes in the auditory condition.

The logarithmic shape of participants' data in the visual-only and auditory-visual sensory conditions is coherent with findings from the literature reporting that the shape of response functions in experiments on numerical estimation of large numerosities generally is negatively accelerated and fits a logarithmic or a power function rather than a linear function [31-33]. The fact that the differences in space occupation are less dramatic between the largest than between the smallest group sizes may play a role in this effect. What is surprising is the linear shape of participants' data in the auditory-only sensory condition, especially as the groups' space occupation is the same in the auditory-only as in the visual-only and auditory-visual sensory conditions thanks to the accurate spatialized audio rendering of our virtual environment.

The different pattern of responses in quantity judgment that we observed for stimuli presented via the auditory channel is consistent with the results of Tokita et al., which exposed different performances in a comparison task of large numerosities with auditory stimuli and with visual stimuli [22]. There also seems to be an influence of sensory presentation of stimuli on approximate numerical judgment with meaningful stimuli, such as crowds of humans. Our findings bring additional support to the idea that complex processes are involved in quantity judgment. Tokita et al. presented the tones and flashes composing their stimuli in a sequential format. Contrastingly, we presented the visual and auditory components of the humanoids composing our stimuli in a simultaneous format. Whereas the higher temporal resolution of auditory processing could explain the findings of Tokita et al., this explanation cannot account for the effect we found with auditory stimuli.

One could also wonder if the effect of the task together with the fixed order of sensory presentation (A then V then AV) and the use of increasing group size presentation can explain the pattern of responses in the auditory condition. The scale that participants had to use to report their quantity judgment was limited from 0 to 10 whereas the number of humanoids composing the groups could potentially increase infinitely. Thus, the differences in the evolution of quantity judgment between sensory modalities could have been linked to a rescaling of participants' responses once they saw the biggest group of humanoids. However, this is unlikely given that in our experiment the responses increased linearly in the first sensory presentation (A) and increased rather logarithmically in the second and third sensory presentation (V and AV) whereas the previous explanation would predict the opposite.

The major difference between our sensory conditions is that, in order to ensure realism, the number of humanoids that were talking was lower than the number of humanoids present in the groups. Given that participants did not underestimate the quantity of humanoids in the auditory condition compared to the visual condition, one possibility is that quantity judgment through the auditory modality involves an implicit arithmetic task: the addition of the silent listeners to the talkers. Whether this implicit arithmetic task takes part in numerical processing or in judgment processes remains to be discovered. We are currently working on an experiment using a more controlled yet less interactive and realistic presentation of crowds in order to further investigate the effect of sensory presentation on numerical estimation with data from a complementary type of experimental setting.

In the auditory-visual condition, participants' responses were better described by a logarithmic rather than a linear function, suggesting that auditory-visual responses are lead by visual quantity judgment. If sensory presentation influences numerical processing, this could be a strategy to preserve cognitive resources and would be consistent with the fact that blind adults are better at magnitude estimates [34] and that their greater accuracy in numerical processing seems to be linked to enhanced high-level cognitive processes [35]. Blind adults would automatically perform quantity estimation with a linear representation of magnitudes.

The linear evolution of quantity judgement that we observe in the auditory condition could also be linked to the characteristics of our stimuli. The auditory component of groups of humans implicates social interaction rules, such as the fact that everyone is not talking at the same time, and thus requests the involvement of an implicit arithmetic task. It is however possible that auditory quantity judgment would be represented logarithmically with meaningful stimuli in which each unit conveys auditory information at the same time (with packs of barking dogs for example). Furthermore, the fact that quantity judgement was not influenced by sensory modality in the bigger groups of humanoids could be explained by the fact that we have to extrapolate for the hidden humanoids when judging the size of crowds through visual modality as well as through the auditory modality. Even if the findings of the present study suggest that quantity judgment is influenced by the sensory modality of meaningful stimuli, further investigation is needed to test whether this effect is found generally in all types of realistic situations.

In contrast to quantity judgments, ratings of discomfort in response to each stimulus were not influenced by the sensory presentation. However, given the influence of sensory presentation on quantity judgement, we can speculate that, at equal judged quantity of humanoids, ratings of discomfort in response to groups of 8 to 48 virtual humans would be modulated by the sensory presentation of the stimulation. Participants' ratings of discomfort increased with stimuli numerosity and rather linearly than logarithmically. Thus, even participants who are not phobic of crowds and not even sensitive to this fear, experienced discomfort when imagining going through our groups of virtual humans.

If different mechanisms are involved in the judgment of quantity of humans according to sensory presentation, it would be interesting to use an auditory-visual virtual environment to explore whether patients with fear of crowds differently judge quantity compared to non-phobic individuals and whether this problem with quantity judgments would be specific to one type of sensory condition and could be depending on an individual's preferred sensory modality (auditory vs. visual). This sensory condition could be then specifically targeted during treatment, increasing the likelihood of obtaining significant exposure therapies for each patient.

References

1. Fyer AJ, Mannuzza S, Martin LY, Gallops MS, Endicott J, Schleyer B, Gorman JM, Liebowitz MR, Klein DF. Reliability of anxiety assessment. II. Symptom agreement. Arch Gen Psychiatry 1989; 46(12):1102-10.

2. Mazumder R, Lambert WE, Nguyen T, Bourdette DN, Cameron MH. Fear of Falling Is Associated with Recurrent Falls in People with Multiple Sclerosis: A Longitudinal Cohort Study. International Journal of MS Care. 2015;17(4):164-170. doi:10.7224/1537-2073.2014-042.

3. Botella C, Villa Martin H, García-Palacios A, Baños RM, Perpiñá C, Alcañiz M. Clinically significant virtual environments for the treatment of panic disorder and agoraphobia. Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society 2004; 7(5): 527–535.

4. Botella C, Villa Martin H, Baños RM, Quero S, Alcañiz M, Riva G. Virtual Reality Exposure in the Treatment of Panic Disorder and Agoraphobia : A Controlled Study . Clinical Psychology & Psychotherapy 2007; 14: 164–175. http://doi.org/10.1002/cpp

5. Malbos E, Rapee RM, Kavakli M. A controlled study of agoraphobia and the independent effect of virtual reality exposure therapy. Australian and New Zealand Journal of Psychiatry 2012; 47(2): 160–168. http://doi.org/10.1177/0004867412453626

6. Moore K, Wiederhold BK, Wiederhold MD, Riva G. Panic and agoraphobia in a virtual world. Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society 2002; 5(3): 197–202.

7. Pelissolo A, Zaoui M, Aguayo G, Yao SN, Roche S, Ecochard R, … Cottraux J. Virtual reality exposure therapy versus cognitive behavior therapy for panic disorder with agoraphobia: A randomized comparison study. Journal of CyberTherapy and Rehabilitation 2012; 5(1): 35–43.

8. Villa Martin H, Botella C, Garcia-Palacios A, Osma J. Virtual Reality Exposure in the Treatment of Panic Disorder With Agoraphobia: A Case Study. Cognitive And Behavioral Practice 2007; 14(1): 58–69. http://doi.org/10.1016/j.cbpra.2006.01.008

9. Vincelli F, Anolli L, Bouchard S, Wiederhold BK, Zurloni V, Riva G. Experiential cognitive therapy in the treatment of panic disorders with agoraphobia: a controlled study. Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society 2003; 6(3): 321–8. http://doi.org/10.1089/109493103322011632

10. Klinger E, Bouchard S, Légeron P, Roy S, Lauer F, Chemin I, Nugues P. Virtual reality therapy versus cognitive behavior therapy for social phobia: a preliminary controlled study. Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society 2005; 8(1): 76–88. http://doi.org/10.1089/cpb.2005.8.76

11. Pertaub DP, Slater M, Barker C. An experiment on public speaking anxiety in response to three different types of virtual audience. Presence Teleoperators and Virtual Environments 2002; 11(1): 68–78. http://doi.org/10.1162/105474602317343668

12. Slater M, Pertaub DP, Steed A. Public Speaking in Virtual Reality: Facing an Audience of Avatars. IEEE Computer Graphics and Applications 1999; 19(2): 6–9. http://doi.org/10.1109/38.749116

13. Slater M, Pertaub DP, Barker C, Clark, DM. An experimental study on fear of public speaking using a virtual environment. Cyberpsychology & Behavior : The Impact of the Internet, Multimedia and Virtual Reality on Behavior and Society 2006; 9(5): 627–33. http://doi.org/10.1089/cpb.2006.9.627

14. Arrighi R, Togoli I, Burr DC. A generalized sense of number. Proceedings of the Royal Society B 2014; 281:20141791.

15. Burr D, Ross J. (2008). A Visual Sense of Number. Current Biology 2008; 18(6): 425–428.

16. Butterworth B. Numerosity Perception: How Many Speckles on the Hen? Current Biology 2008; 18(9): R388–R389.

17. Feigenson L, Dehaene S, Spelke E. Core systems of number. Science 2004; 8(7): 307–14. http://doi.org/10.1016/j.tics.2004.05.002

18. Gallistel CR, Gelman R. Non-verbal numerical cognition: From reals to integers. Trends in Cognitive Sciences 2000; 4(2): 59–65.

19. Whalen J, Gallistel CR, Gelman R. Nonverbal Counting in Humans: The Psychophysics of Number Representation. Psychological Science 1999; 10(2): 130-137.

20. Wynn K. Psychological foundations of number: Numerical competence in human infants. Trends in Cognitive Sciences 1998; 2(8): 296–303.

21. Barth H, Kanwisher N, Spelke E. The construction of large number representations in adults. Cognition 2003; 86: 201–221.

22. Tokita M, Ashitani Y, Ishiguchi A. Is approximate numerical judgment truly modality-independent? Visual, auditory, and cross-modal comparisons. Attention, Perception & Psychophysics 2013; 75(8): 1852–61.

23. Moeck T, Bonneel N, Tsingos N, Drettakis G, Viaud-Delmon I, Alloza D. 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.

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

25. O'Sullivan, C., & Ennis, C. (2011, May). Metropolis: multisensory simulation of a populated city. In *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2011 Third International Conference on* (pp. 1-7). IEEE.

26. Spielberger CD, Gorsuch RL, Lushene PR, Vagg PR, Jacobs AG. (1983) Manual for the State-Trait Anxiety Inventory (Form Y). Palo Alto, CA: Consulting Psychologists Press.

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

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

29. Bouchard S, Côté S, St-Jacques J, Robillard G, Renaud P. Effectiveness of virtual reality exposure in the treatment of arachnophobia using 3D games. Technology and Health Care : Official Journal of the European Society for Engineering and Medicine 2006; 14(1): 19–27.

30. Rothbaum BO, Hodges LF, Kooper R, Opdyke D, Williford JS, North MM. Virtual Reality Graded Exposure in the Treatment of Acrophobia : A Case Report. Behavior Therapy 1995; 26: 547–554.

31. Banks WP, Coleman, MJ. Two subjective scales of number. Perception & Psychophysics 1981; 29(2): 95–105.

32. Dehaene S, Izard V, Spelke E, Pica P. Log or linear ? Distinct intuitions of the number scale in Western and Amazonian indigene cultures. Science 2008; 320(5880): 1217–1220. http://doi.org/10.1126/science.1156540.

33. Izard V, Dehaene S. Calibrating the mental number line. Cognition 2008; 106(3): 1221–1247.

34. Castronovo J, Seron X. Numerical estimation in blind subjects: Evidence for the impact of blindness and its following experience. Journal of Experimental Psychology. Human Perception and Performance 2007; 33(5): 1089–1106.

35. Castronovo J, Delvenne JF. Superior numerical abilities following early visual deprivation. Cortex 2013; 49(5): 1435–1440.