2017 IEEE World Haptics Conference (WHC)
Fürstenfeldbruck (Munich), Germany
6–9 June 2017

How Many Objects are Inside this Box?*

Myrthe A. Plaisier and Jeroen B.J. Smeets

Abstract— **To judge the contents of a box, we do not necessarily have to open it. By shaking a box we can make an estimate of its contents based on haptic and auditory information. Not much is known about the perception of properties of objects that are inside a box. In this study we investigated how accurately participants can judge the number of wooden spheres inside a small handheld box by shaking the box. This was done in a 'haptic + auditory' condition in which participants shook the box and in a subsequent 'auditory only' condition in which recorded sounds from the trials in the haptic + auditory condition were played back. In both conditions participants had to judge the number of spheres (1 to 5) inside the box. In the haptic + auditory condition participants could perform this task accurately for up to about 3 spheres, while for larger numbers they systematically underestimated the numerosity. Although participants could perform this task above chance in both conditions, accuracy was lower in the auditory condition than in the haptic + auditory condition. By actively shaking the box the number of objects inside can be judged accurately for up to 3 objects.**

I. Introduction

Since the introduction of handheld devices such as cell phones and tablets the possibilities of using haptic feedback in these devices have been investigated. Over the past years, the idea of using gestures that involve shaking the device has been explored [1, 2]. This resembles a situation in which we shake a box to judge its contents, for instance, to know how much cereal is left in a cereal box we can shake the box to can make an estimate of how much is left. This would not allow someone to determine the exact number of remaining cornflakes, but it is good enough to know whether there is enough cereal left for breakfast. However, for small numbers of objects it might actually be possible to estimate the exact number of objects inside a box by shaking. Here we investigated whether this is indeed the case.

It has been shown that humans can haptically judge the number of objects held in the hand or pressed to the fingers quickly and accurately [3-6]. In those studies, however, there was always direct contact between the objects/stimuli and the hand. To accurately judge the exact number of objects inside a box, the number of unique objects needs to be deduced from hearing or feeling consequences of the movements of the objects. This is a more complex task than judging the number of objects held in the hand. The cues for the number of objects are the number of collisions with the wall or between objects in a certain time interval, combined with information about the rolling movements of the objects. The number of collisions and the vibrations related to the rolling will increase with the number of objects, but will depend also on the size of the objects relative to that of the box and the speed and frequency of the shaking movements.

Aside from haptic information about collisions with the sides of the box and between objects, it might also be possible to judge the number of items from the sounds that are generated during collisions. Numerosity judgement in audition has been investigated mostly with sequences of sounds (for instance [7, 8]). In that case the number of beeps in the sequence has to be judged. Humans can do this accurately, but again only for small numbers. When shaking a box, however, the task would be to reconstruct the number of objects responsible for the sounds made by the collisions, not the number of collisions itself.

Judging the number of objects inside a box could be interpreted as a sort of multiple object-tracking task. Once a collision is assigned to a unique object it is important to keep track of that object to know whether another collision is likely to be caused by that same object or whether there is a second object inside the box. That way the exact number of objects inside the box can be determined. From research on visual tracking of multiple objects it has suggested that humans can keep track of five different objects simultaneously [9]. We therefore expect that judging the number of objects inside a box will only possible for a small set of objects, if possible at all.

Here we investigated whether participants could judge the exact number of objects (in this case spheres) inside a handheld box by shaking the box ('haptic + auditory condition'). To obtain an insight in the importance of actively shaking the box we added a condition in which participants only heard a playback of the sounds that were recorded during the auditory + haptic trials in which they actively shook the box ('auditory only' conditions).

II. Methods

A. Participants

Seven members of the department of Human Movement Sciences at Vrije Universiteit Amsterdam volunteered to participate in the experiment (age range 23 to 34 years). All were self-reported right handed and were naive as to the purpose of the experiment. This experiment was part of a project that was approved by the ethical committee of the department of Human Movement Sciences at Vrije Universiteit Amsterdam.

B. Stimuli and Procedure

The stimulus consisted of a cardboard box (7 x 6.5 x 3.5 cm). Inside this box there could be one, two, three, four, or five wooden spheres. These could be either all small spheres (1 cm diameter, 1 g) or all large spheres (2 cm diameter, 3 g). The sphere size was varied to make the weight of the box an unreliable cue for the number of spheres inside. The box weighed 9 grams. Prior to the experiment participants were shown the box and the two possible sphere sizes. They were told that the spheres could either be large or small, but that

* This work was financed by a VENI grant (2012/23306/MAGW) from the Dutch Organisation for Scientific Research (NWO).

M.A. Plaisier was with the Department of Human Movement Sciences, Vrije Universiteit Amsterdam. She is now with the Department of Biomechanical Engineering, Delft University of Technology, (e-mail: m.a.plaisier@tudelft.nl).

J.B.J Smeets is with the Department of Human Movement Sciences, Vrije Universiteit Amsterdam (J.B.J.Smeets@vu.nl).

978-1-5090-1424-8/17/$31.00 ©2017 IEEE

they did not have to pay attention to the sphere size. Participants were not told what the maximum number of spheres was and never received feedback on whether their answers were correct. Each number of spheres was presented 10 times for each size, randomly interleaved. This resulted in 100 trials total.

In the 'haptic + auditory' condition, participants sat at a table and placed their dominant hand behind a curtain occluding the hand from view. We did not expect that seeing the box would change the performance, but it was more practical to keep the stimuli out of the field of view of the participant throughout the experiment. They placed their hand with the palm facing upwards on a small platform that was mounted on the space bar of a keyboard. The experimenter placed the box in the hand of the participant (Figure 1). When participants lifted their hand, the release of the space bar was registered and after 5 s an alarm sound was played to indicate the end of the trial. Participants were instructed to return their hand to the platform immediately when they heard the end-of-trial alarm. In this way the exploration time was controlled. From the moment of the release of the space bar, sound was recorded for the entire duration of a trial using the built-in microphone of a Macbook Pro. Participants were instructed to always complete the full 5 s and not to answer before the end of the trial was signaled. After returning their hand, the participants were asked to verbally indicate the number of spheres they thought were inside the box.

The haptic + auditory condition was followed by the 'auditory only' condition. In the auditory only condition participants sat at the table while the experimenter played back the sound recordings from all the 100 trials in the haptic + auditory condition in randomized order. Participants were told that the sounds were recorded from their own previously performed trials. Again they were asked to judge the number of spheres inside the box, now after listening to the full 5 seconds of audio recording.

III. Results

Participants were able to perform the task in both conditions: the reported numerosity increased with the presented number of spheres for each individual participant not only in the haptic + auditory condition (Figure 2a), but also in the audition only condition (Figure 2b). In the haptic + auditory condition participants were accurate for up to 3 spheres and started to systematically underestimate for larger numbers. In the auditory only condition participants show a slightly larger underestimation for larger numbers, and in addition an overestimation for small numbers. Note that these trends were visible for each of the participants individually.

The reported numerosity as a function of the presented numerosity was similar for the small and large spheres (Figure 2c and 2d). This indicates that the weight of the box or other possible cues that correlate with sphere size did not influence the perceived numerosity.

The standard deviation in the answers for the 10 repetitions increased slightly with the number of spheres (Figure 3a). Furthermore, the larger standard deviations in the auditory only condition indicate that participants were more variable in their responses in the auditory only condition than

Figure 1. A participant holding the box with the hand still resting on the platform mounted on the keyboard. Lifting the hand triggered the start of a trial. The hand was occluded from view by a curtain.

in the haptic + auditory condition. These trends were confirmed to be statistically significant with a repeated measures ANOVA with factors condition and numerosity that showed a main effect of condition ($F(1, 6) = 35.6$, $p = 0.001$) and of numerosity ($F(1.7, 10.6) = 14.4$, $p = 0.001$, Greenhouse Geisser correction applied), and no interaction effect.

For each condition the root mean squared deviation with respect to correct performance was calculated (Figure 3b). Here smaller values indicate better performance. Participants performed better in the haptic + auditory condition than in the auditory only condition (paired t-test, $t(6) = 6.8$, $p = 0.0005$).

IV. Discussion

The results show that participants could determine the number of spheres inside a box quite well by shaking. This was especially the case in the haptic + auditory condition. In this condition, participants performed accurately for up to 3 items. For the largest numerosities (4 and 5), a systematic underestimation occurred. In the auditory only condition, a similar underestimation for large numerosities was combined with an overestimation for the smallest numerosities, indicating that there was regression to the mean. A regression to the mean indicates that participants are using information that is not related to the present trial such as guessing or using a prior. The fact that the variability in the judgements was larger in the auditory only condition suggests that the amount of guessing was greater in the auditory only condition.

Why would participants be more uncertain in the auditory only condition than in the haptic + auditory condition? We mentioned in the introduction that the interpretation of the cues requires an estimation of the shaking movement. In the haptic + auditory condition participants had exact information about in which direction and how fast they were shaking the box. Of course, this was not the case in the auditory only condition. Therefore, it was probably difficult to tell whether sounds originated from a single sphere shaken quickly or two (or more) spheres shaken more slowly. This might have led to more guessing and thus have caused the

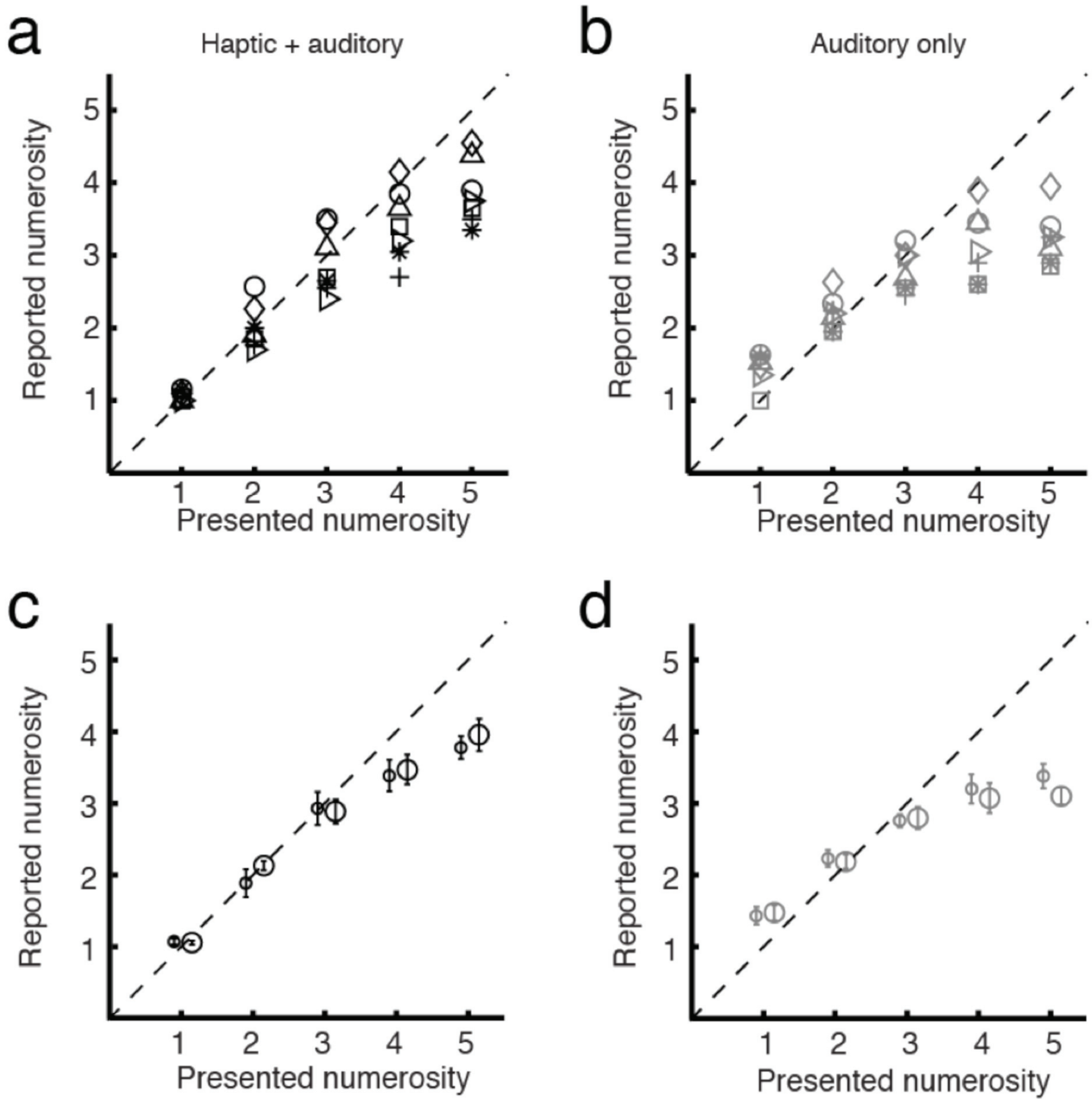

Figure 2. Average numerical response for each presented numerosity in both conditions. The dashed lines indicate veridical performance. (a,b): Data averaged over sphere size, shown for each of the seven participants individually. (c,d): Data averaged over participants for the small (small symbols) and large spheres (large symbols) separately for both conditions, plotted with a small horizontal offset for clarity. Error bars indicate the between-participants standard error of the mean

observed regression to the mean. Also, in the auditory only condition participants heard a recording while they heard live sounds during the haptic + auditory condition. The recording is of course of somewhat poorer quality than the live sounds.

Auditory numerosity judgment studies usually require participants to judge the number of beeps in a sequence. The current study, however, is more complicated, as the task was not to judge the total number of sounds. Instead, the number of unique objects responsible for generating the sounds had to be reconstructed. Although this was more difficult based on sound alone than when actively shaking the box, participants still performed better than chance. So this task is doable based on sound alone. Here it must be noted, however, that in our study participants always first performed the haptic + auditory condition and they were aware that their own trials were used in the auditory condition. This means they had some idea about, for instance, how quickly they had been shaking the box during the haptic + auditory condition and they had gotten used to the type of sounds that were generated. The task in the auditory only condition might have been much more difficult if participants would not have performed the haptic + auditory condition first. Also, the results might be different if the recordings from a different participant were presented. So, performance in the auditory only condition in our experiment might not generalize to other situations.

Using your memory of how you moved in the haptic + auditory condition provides another explanation of the

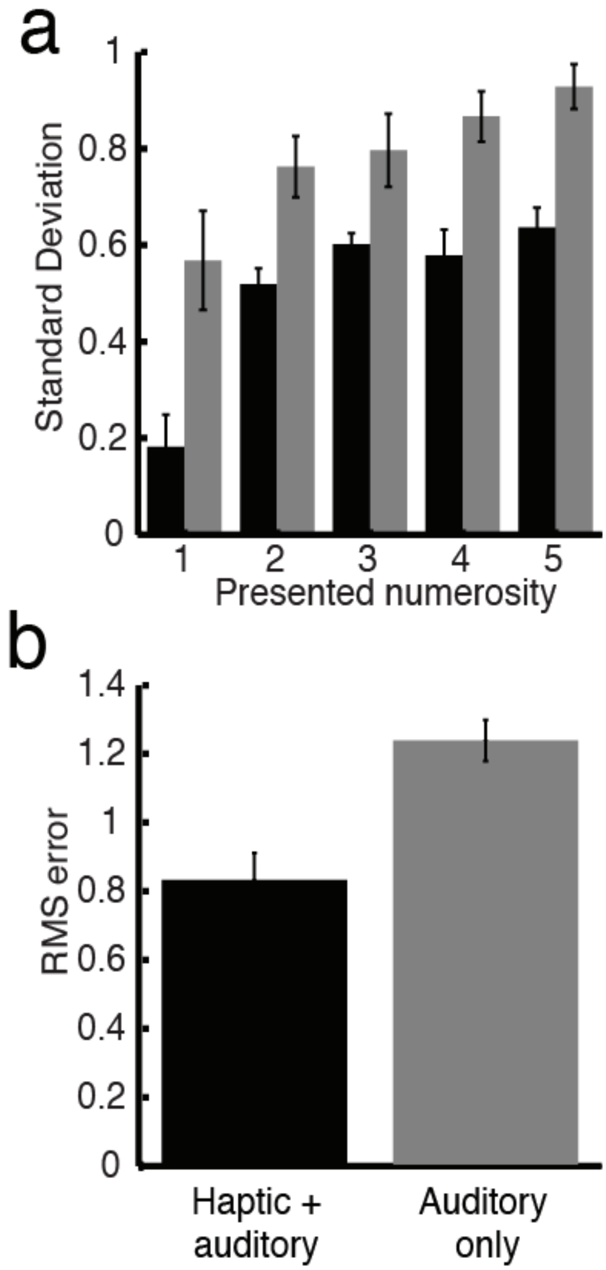

Figure 3. (a) Within-participant standard deviation of the responses averaged over participants for each numerosity in both conditions. (b) The root mean squared deviation from correct performance averaged over participants for both conditions. Error bars indicate standard error of the mean.

difference in results between the two conditions, unrelated to knowledge of the active movement: participants might have used a consistent shaking strategy across trials, so knowledge of the shaking movement was equally present in the auditory only condition. The cause of the different results might be that in the auditory only condition, participants had unimodal information movements and collisions of the spheres, whereas the information was bimodal (auditory and tactile) in the haptic + auditory condition. Combining information from two modalities leads to more precise estimates [e.g. 10, 11], and therefore reduces the reliance on prior information [e.g. 12]. These two aspects explain both the lower variability and reduced regression to the mean in the haptic + auditory condition.

To perform either of the conditions accurately the participants would have to reconstruct the number of unique spheres causing the collisions with the walls of the box as well as collisions between spheres. Alternatively, participants could make an estimate by simply judging the number of collisions per time interval. Estimation is less accurate and less precise than counting. In the haptic + auditory condition participants were highly accurate for small numbers of spheres, which suggests that participants indeed reconstructed the number of spheres from the collisions. This requires some sort of internal representation of the dynamics of the objects inside the box. It has been shown that humans do have implicit knowledge of physical relationships [13, 14]. This can be used even in virtual settings in which information is usually poorer. It has been shown that participants can perform implicit tasks like estimating tube length by letting a ball role through it even in a virtual setting [15]. Furthermore, participants can, for instance, judge the mass, position or friction of virtual objects inside a box [16]. The current results show that that participants could use such cues to accurately judge the number of objects inside a box for up to 3 objects in case of actively shaking the box.

In recent years different types of handheld devices are being developed that present virtual contents rolling or rattling inside the device [1, 17-19]. For the design of such virtual contents it is useful to take perceptual limits into account. As shown in the present study, the number of objects inside a box can be accurately judged, but only for up to about 3 items.

References

[1] J. Williamson, R. Murray-Smith, and S. Hughes, "Shoogle: Excitatory Multimodal Interaction on Mobile Devices," in *Proc. CHI 2007*, pp. 121-124, 2007.

[2] J. Linjama and T. Kaaresoja, "Novel, minimalist haptic gesture interaction for mobile devices," in *Proc. CHI 2004*, pp. 457-460, 2004

[3] M. A. Plaisier, W. M. Bergmann Tiest, and A. M. L. Kappers, "One, two, three, many - Subitizing in active touch," *Acta Psychologica,* vol. 131, pp. 163-170, 2009.

[4] K. J. Riggs, L. Ferrand, D. Lancelin, L. Fryziel, G. Dumur, and A. Simpson, "Subitizing in tactile perception," *Psychological Science,* vol. 17, pp. 271-272, 2006.

[5] M. A. Plaisier, W. M. Bergmann Tiest, and A. M. L. Kappers, "Grabbing subitizing with both hands: bimanual number processing," *Experimental Brain Research,* vol. 202, pp. 507-512, 2010.

[6] Z. Z. Cohen and A. Henik, "Effects of Numerosity Range on Tactile and Visual Enumeration," *Perception,* vol. 45, pp. 83-98, 2016.

[7] V. Camos and B. Tillmann, "Discontinuity in the enumeration of sequentially presented auditory and visual stimuli," *Cognition,* vol. 107, pp. 1135-1143, 2008.

[8] M. Tokita and A. Ishiguchi, "Precision and Bias in Approximate Numerical Judgment in Auditory, Tactile, and Cross-modal Presentation," *Perception,* vol. 45, pp. 56-70, 2016.

[9] Z. W. Pylyshyn, "Visual indexes, preconceptual objects, and situated vision," *Cognition,* vol. 80, pp. 127-158, 2001.

[10] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature,* vol. 415, pp. 429-433, 2002.

[11] R. J. van Beers, A. C. Sittig, and J. J. D. van der Gon, "Integration of proprioceptive and visual position-information: An experimentally supported model," *Journal of Neurophysiology,* vol. 81, pp. 1355-1364, 1999.

[12] D. Kersten, P. Mamassian, and A. Yuille, "Object perception as Bayesian inference," *Annu Rev Psychol,* vol. 55, pp. 271-304, 2004.

[13] T. L. Hubbard, "Environmental Invariants In The Representation Of Motion - Implied Dynamics And Representational Momentum, Gravity, Friction, And Centripetal Force," *Psychonomic Bulletin & Review,* vol. 2, pp. 322-338, 1995.

[14] J. McIntyre, M. Zago, A. Berthoz, and F. Lacquaniti, "Does the brain model Newton's laws?," *Nature Neuroscience,* vol. 4, pp. 693-694, 2001.

[15] H. Yao and V. Hayward, "An Experiment on length perception with a virtual rolling stone," in *Proceedings of the Eurohaptics Conference*, 2006, pp. 325-330.

[16] Y. Tanaka and K. Hirota, "Shaking a Box to Estimate the Property of Content," in *LNCS* vol. 7282, P. Isokoski and J. Springare, Eds., Springer-Verlag Berlin, pp. 564-576, 2012.

[17] Y. Sekiguchi, K. Hirota, and M. Hirose, "Haptic Interface Using “Estimation of Box Contents” Metaphor," in *ICAT*, Tokyo, 2003.

[18] T. Yamamoto and K. Hirota, "Device for Estimation of Weight and Material of Contents by Shaking," in *Proceedings of the Eurohaptics Conference*, Paris, 2014, pp. 403-410.

[19] Y. Sekiguchi, S. Matsuoka, and K. Hirota, "Inertial Force Display to Represent Content Inside the Box," in *Proceeding of the Eurohaptics Conference*, Amsterdam, 2010, pp. 81-86.