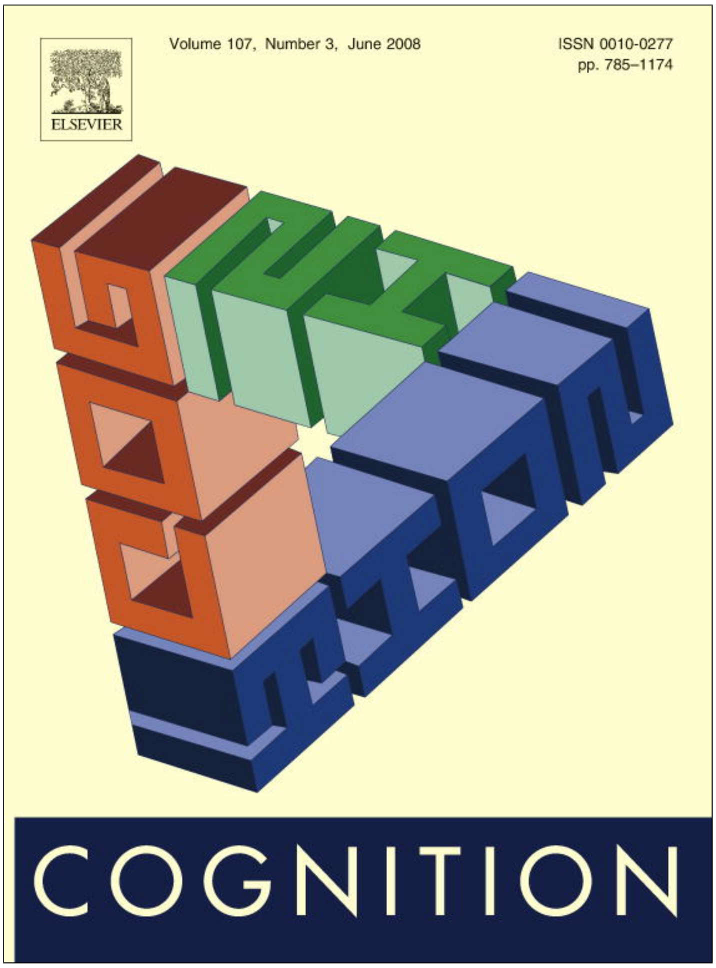
Volume 107, Number 3, June 2008
ISSN 0010-0277
pp. 785–1174
ELSEVIER
COGNITION

Available online at www.sciencedirect.com
ScienceDirect
Cognition 107 (2008) 1135–1143
COGNITION
www.elsevier.com/locate/COGNIT

Brief article

Discontinuity in the enumeration of sequentially presented auditory and visual stimuli

Valérie Camos [a,*], Barbara Tillmann [b]

[a] *Université de Bourgogne & Institut Universitaire de France, LEAD-CNRS, Pôle AAFE, Esplanade Erasme, B.P. 26513, 21065 Dijon Cedex, France*
[b] *Université Claude Bernard Lyon 1, Université de Lyon, Neurosciences Sensorielles, Comportement, Cognition, CNRS-UMR 5020, IFR 19, Lyon, France*

Received 4 December 2006; revised 19 October 2007; accepted 3 November 2007

Abstract

The seeking of discontinuity in enumeration was recently renewed because Cowan [Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–185; Cowan, N. (2005). *Working memory capacity*. Hove: Psychology Press] suggested that it allows evaluating the limit of the focus of attention, currently estimated at four items. A strong argument in favour of a general constraint of the cognitive system is that similar discontinuities should be observed in modalities different from the classic simultaneous presentation of visual objects. Recently, data were provided on tactile stimuli, but the authors diverged in their conclusion about the existence of such discontinuity [Gallace, A., Tan, H. Z., & Spence, C. (2006). Numerosity judgments for tactile stimuli distributed over the body surface. *Perception, 35*(2), 247–266; Riggs, K. J., Ferrand, L., Lancelin, D., Fryziel, L., Dumur, G., & Simpson, A. (2006). Subitizing in tactile perception. *Psychological Science, 17*(4), 271–272]. Following a similar rationale, our study aimed at evaluating discontinuity in the enumeration of auditory and visual stimuli presented sequentially. The clear and similar discontinuity observed in error rates, response times and given responses for both modalities favours the general capacity limit view, but also questions the size of this capacity, because the discontinuity occurred here at size 2. However, the masking of stimuli in sensory memory could not be entirely discarded.

* Corresponding author.
E-mail address: valerie.camos@u-bourgogne.fr (V. Camos).

doi:10.1016/j.cognition.2007.11.002

Keywords: Focus of attention; Subitizing; Counting; Sequential presentation; Auditory stimuli

1. Introduction

Since more than a century, studies on enumeration of visual objects have shown a discontinuity between fast and accurate performance for small collections, whereas larger arrays are slowly and erroneously quantified (Jevons, 1871). In the latter, the serial counting of objects lead to a linear increase of error rates and response times (RTs) with size, whereas in the former (referred to as the subitizing range), no such increase is observed and the slope of RTs being very small (around 40 ms/item; Kaufman, Lord, Reese, & Volkmann, 1949; Trick & Pylyshyn, 1994).

Recently, Cowan (2001,2005) suggested that the capacity limit of the focus of attention could be estimated by examining performance discontinuities such as in enumeration. The lack of increase of error rates and RTs with the number of objects would reveal the number of objects that can be conjointly held within the focus of attention. When this number of objects overcomes the capacity limit, the focus of attention needs to be moved sequentially among the objects, which induces an increase of error rates and RTs. This theoretical view predicts similar discontinuity patterns in modalities other than the simultaneously presented visual stimuli used in previous enumeration studies (Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). Recently, two papers evaluated enumeration in tactile perception (Gallace, Tan, & Spence, 2006; Riggs et al., 2006), but they led to opposite conclusions about the existence of a discontinuity in this modality. The former observed a linear relationship between the number of tactile stimulations and both mean RTs and error rates, whereas the latter showed a discontinuity between 1–3 and 4–6 stimulations on accuracy and RTs. It might be that the rather uncommon use of tactile information to evaluate numerosity obscured the results. Thus, our study tested the auditory modality, in which numerosity processing is used more frequently (e.g., the number of sounds of bells indicates the time; Garner, 1951; John, 1972; Massaro, 1976).

Numerosity processing for auditory material can concern both simultaneous and sequential presentations of sounds. However, simultaneous processing of sounds requires processes of fusion and stream segregation (Bregman, 1990), and these processes might veil the perceived number of separated sources. To avoid acoustical ambiguity, we adapted a sequential stimulus presentation in the auditory modality and, for sake of comparison, in the visual modality. However, to prevent the sequential processing of the stimuli, i.e., their counting, they were presented rapidly (Stimulus Onset Asynchrony = 80 ms); the enumeration of the French numberline taking around 160 ms per item (Camos, Barrouillet, & Fayol, 2001). Albeit this sequential presentation, the stimuli have to be simultaneously maintained in the focus of attention to evaluate their numerosity. As the capacity limit of the focus of attention is not exceeded, the representations of the stimuli would not suffer from a time decay of their activation and could then be enumerated. In our present study, the auditory stimuli were pure tones at different pitch heights without tonal relationships and the visual stimuli were colour dots presented sequentially at the centre of the screen. These material and presentation forms excluded any pattern recognition, a mechanism that is often mentioned to account for the performance observed in small size collections (Mandler & Shebo, 1982; Peterson & Simon, 2000). Our study aimed at seeking for the discontinuity in the enumeration of auditory and visual modalities presented sequentially. The hypothesis of a domain-general limit of processing predicts a similar discontinuity in performance for both modalities. However, if enumeration relies even partly on some modality-specific processes, its limit should differ between the two modalities.

2. Methods

2.1. Participants

Twenty undergraduate students (mean age = 22.5 years; *SD* = 2.5 years; 7 males) of the Université René Descartes – Paris V participated in the experiment for partial course credit. They had never participated to any enumeration study before, and none of them reported having hearing or sight difficulties.

2.2. Material

Sixty sequences of one to six events were created for both auditory and visual material. The duration of presentation of each tone or dot was 60 ms with an Inter-Stimulus-Interval of 20 ms. To prevent potential influences of forward masking between tones, each tone within a sequence had a different pitch height (i.e., at 200, 250, 300, 350, 400, and 450 Hz). These tones were not tuned to the pitch height of chromatic tones in the Western tonal system. For the visual presentation, within a sequence, each 1.5-cm dot had a different colour to mirror the changes between the tones in the auditory sequences.

2.3. Procedure

Participants evaluated the number of stimuli in a sequence presented over headphones or on the centre of a white screen. Half of them started with the visual modality. Within each trial, the sequence was presented after a 500-ms visual ready signal. At the end of the sequence, the participant's oral response triggered the voice key. The trials were presented in a fixed random order different for the two modalities. For each modality, training consisted of six sequences for which the experimenter gave error feedback.

3. Results

On average 1% of the data were discarded due to problems with the voice key. This percentage did not differ between conditions. It should be noted that data analyses of response times and of correct response times led to same result patterns with similar significant effects. We chose to report analyses on response times, because correct response time analyses were performed for a smaller number of participants due to missing data, especially in some conditions (e.g., for large set sizes).[1]

To evaluate whether participants were able to simply detect the used stimuli, we first analysed percentages of errors when only one item was presented. Five participants almost systematically reported seeing two dots when only one was presented (between 90% and 100% of the trials). These participants were discarded from the analyses, because we cannot preclude the cause of their perceptual difficulty. Although all self-reported no hearing or sight problems, these participants may have a longer visible persistence that could have affected their perception. They belonged to both groups, with three of them starting with the visual modality and two of them starting with the auditory modality. Because the order of presentation of the modalities did not affect percentages of error, response times, and responses given, and did not interact with Size and Modality, this factor was not included in the following analyses.

First, to evaluate the discontinuity of enumeration performance, we compared linear and logarithmic models to a bilinear model in predicting error rates and RTs in both modalities. For auditory stimuli, the bilinear model ($R^2 = .73$ and $R^2 = .72$, $p\text{s} < .0001$) was a better account of errors rates and RTs, respectively, than the linear ($R^2 = .48$ and $R^2 = .35$, $p\text{s} < .0001$) and the logarithmic ($R^2 = .12$, $p = .01$, and $R^2 = .42$, $p < .0001$) models. Similarly for visual stimuli, the bilinear model ($R^2 = .77$ and $R^2 = .55$, $p\text{s} < .0001$) outperformed the linear ($R^2 = .40$ and $R^2 = .18$, $p\text{s} < .0001$) and the logarithmic ($R^2 = .15$, $p < .01$, and $R^2 = .22$, $p < .0001$) models in predicting error rates and RTs, respectively.

As these first analyses showed a discontinuity in enumeration performance, we performed a series of paired-sample t-tests comparing size n to size $n + 1$ for percentages of errors and response times to evaluate the changing point. For both percentages of error and response times, size 2 was the smaller size for which the comparisons were significant, $p\text{s} < .01$. The same analyses were performed independently for the auditory and the visual modality, and led to the same limit,

[1] The analyses on correct response times were restricted to 10 participants because 5 participants had 100% error in size 5 or/and 6 in either visual or auditory modality. Nevertheless, all effects observed for response times were replicated with correct response times. Indeed, the bilinear model ($R^2 = .68$ and $R^2 = .61$, $p\text{s} < .0001$) was a better account of the performance in the visual and the auditory modality, respectively, than the linear ($R^2 = .18$, $p < .001$, and $R^2 = .27$, $p < .0001$) and logarithmic models ($R^2 = .18$, $p < .001$, and $R^2 = .35$, $p < .0001$). Moreover, the two ANOVAs with Modality and Size as within-participant factors on size up to 2 and on sizes 3–6, respectively, revealed the same effects. In the former, only the modality effect was significant, $F(1,9) = 7.00$, $p = .03$; the size and interaction effects were non-significant, $p\text{s} > .10$. In the latter, only the size effect was significant, $F(3,27) = 7.75$, $p < .001$; the modality and interaction effects were non-significant, $p\text{s} > .18$.

ps < .05. Thus, analyses of variance (ANOVAs) were performed separately for the subitizing (up to 2) and the counting (sizes 3–6) ranges on percentages of error and response times with Modality and Size as within-participant factors. For the subitizing range, the visual modality led to higher percentages of errors and longer response times than the auditory modality (Table 1), $F(1,14) = 8.77$, $p = .01$ and $F(1,14) = 10.39$, $p < .01$, respectively. As the t-tests showed, Size had no significant effect, Fs < 1, and did not interact with Modality, Fs < 1. For the counting range, as expected, percentages of errors and response times increased significantly with Size, $F(3,42) = 23.25$, $p < .0001$ and $F(3,42) = 11.71$, $p < .0001$, respectively. Although higher percentages of errors were committed in the visual modality, $F(1,14) = 6.17$, $p = .03$, the two modalities did not differ on response times, $F(1,14) = 2.84$, $p = .11$. The interactions between Size and Modality were never significant, Fs < 1. To evaluate the slope in the subitizing and in the counting range, regression analyses were performed on response times for both visual and auditory modalities. As classically found in the literature (see Trick & Pylyshyn, 1994, for a review), the subitizing slope was around 40 ms (visual = 32 ms and auditory = 37 ms), and the counting slope much higher (visual = 261 ms and auditory = 374 ms). Although the subitizing slopes and the counting slopes did not differ significantly between modalities, ts < 1, the counting slope was significantly greater than the subitizing slope for auditory stimuli, $t(14) = 4.53$, $p < .001$, and marginally significantly greater for visual stimuli, $t(14) = 1.83$, $p = .09$.

Finally, analyses of the responses given provided further evidence for a dissociation between subitizing and counting ranges. In the counting range, the difference between the given and the correct response increased linearly with the size for both visual and auditory modalities, $F(3,42) = 19.37$, $p < .0001$ and $F(3,42) = 3.61$, $p = .02$, respectively. In this range, participants significantly underestimated the size, $F(1,14) = 10.59$, $p < .01$ and $F(1,14) = 7.10$, $p = .02$ for visual and auditory modalities, respectively. In the subitizing range, there was no difference between the given and the correct response in the auditory modality, $p > .30$, but participants overestimated the size in the visual modality, $F(1,14) = 9.26$, $p < .01$.

Table 1
Mean percentages of errors, response times (ms), and responses given (and *SD*) according to the number of items and the modality

Modality		Number of stimuli					
		1	2	3	4	5	6
Auditory	Error	0 (0)	1 (5)	15 (18)	25 (22)	38 (27)	53 (28)
	RT	741 (115)	778 (231)	1278 (470)	1622 (732)	2120 (1151)	2360 (1455)
	Response	1.00 (0)	1.99 (0.05)	2.91 (0.16)	3.80 (0.27)	4.79 (0.47)	5.60 (0.56)
Visual	Error	14 (20)	17 (24)	32 (23)	49 (26)	63 (25)	63 (28)
	RT	924 (259)	957 (315)	1129 (322)	1408 (627)	1546 (792)	1954 (1453)
	Response	1.14 (0.21)	2.15 (0.24)	3.01 (0.35)	3.79 (0.46)	4.51 (0.47)	5.27 (0.64)

4. Discussion

The aim of the present study was to evaluate the discontinuity in performance in the enumeration of auditory and visual stimuli presented sequentially. The results support the existence of such discontinuity in both modalities. For the smaller collections (i.e., one and two items), the RTs were short and the error rates were low and both did not differ with the size of the collections. However, for the larger collections, both RTs and error rates increased linearly with the number of items. This pattern of results is akin to what is classically described in the enumeration of simultaneously presented visual stimuli (Mandler & Shebo, 1982; Trick & Pylyshyn, 1994). Thus, it supports the idea of a capacity-limited process similar for both auditory and visual modalities, even when stimuli are presented sequentially. Moreover, this process could not rely on pattern recognition because presentation was sequential on the center of the screen and without tonal relationships between the sounds.

Two theoretical accounts could explain our results. First, according to Cowan (2001, 2005), such discontinuity reflects a general attention capacity, namely the limit of the focus of attention. All items that can be maintained in a single focalisation can be processed at once and their quantity be evaluated. Second, in the FINSTs (Fingers of INSTanciation) model, Trick and Pylyshyn (1994) proposed that enumeration is sustained by the assignation of individuation pointers, and a preattentive rapid mechanism allows determining the number of assigned FINSTs without counting. This assignment would thus not be restricted to spatially displayed objects, as in the original model, but could be extended to any types of stimuli and presentations, as Riggs et al. (2006) have suggested for tactile perception. Although our study was not designed to disentangle between these two accounts, the results showed that the enumeration process relies on a domain-general mechanism (i.e., at least for auditory and visual modalities).

Moreover, albeit their contrasting positions on the attentional demand, the two accounts propose the same limit. In Cowan's model, the focus of attention is limited to four chunks and in Trick and Pylyshyn's model, the maximum number of FINSTs is also 4. In comparison to the previously used simultaneous presentation of visual items, we observed a reduction in the amount of information that could be processed (i.e., down to 2). Such a small limit is not so uncommon, and a similar limit has been reported in the subitizing range for simultanagnosic adults, patients suffering from visual-attentional deficit (Dehaene & Cohen, 1994), for adolescents suffering from autism (Gagnon, Mottron, Bherer, & Joanette, 2004), for children suffering from mental retardation (Lépine, 2003), and for young children (Gelman & Tucker, 1975; Lépine, 2003). Two interpretations could be proposed to understand this reduced limit. The simplest account of our results is that the size of the focus of attention or the number of FINSTs is indeed limited at 2. Thus, previous studies, in which stimuli were simultaneously presented, and often for an unlimited duration, overestimated this limit because participants could focus twice on the display. In favour of this interpretation, it should be noted that studies on enumeration showed an increase of RTs even within the subitizing range (see Dehaene & Cohen, 1994, for a review).

Alternatively, the rapid presentation rate we used could induce an underestimation of the capacity limit. Within the sensory memory, in which the representations of the stimuli are stored before being processed, the representations could suffer from masking. As Massaro (1975) showed, auditory stimuli suffer from backward masking and a sound could interfere with the storage of an earlier sound when presented rapidly (for a similar phenomenon in vision, see Turvey, 1973). Thus, the focus of attention would attend to impaired representations, which would directly affect the efficiency of further processes. Our results so far cannot disentangle between these two interpretations. However, it is important to note that Massaro (1975) showed backward masking for pitch height discrimination (and not detection itself) and that pitch height processing was irrelevant for the enumeration task requested in our study. To investigate whether masking was nevertheless involved in our result pattern (and notably for central rather than peripheral processes), future studies should increase the used SOAs (while still remaining beyond SOAs allowing for counting). If the second interpretation is correct, the masking should reduce, and the discontinuity should occur at a larger set size. Although for a different purpose, Repp (2007) has recently tested the impact of the SOA on the enumeration of serially presented auditory events. Interestingly, his results give strong support to a reduced limit of the subitizing of the auditory stimuli. He estimated this limit at 2–3 events, a rather similar limit than the one we observed here. Furthermore, increasing the SOA from 80 to 170 ms by steps of 10 ms (i.e., still beyond the speed of verbal counting) did not affect the percentage of correct answer, or the deviation from correct counts. Although this increase of SOA reduced the RTs, the interaction with the size of the presented set of stimuli was not significant, and a similar limit occurred for the 10 different SOA conditions. Although these results brought argument against masking phenomena in sensory memory to explain this reduced capacity limit, Repp (2007) did not compare across modalities and his data do not allow to reject the possible influence of SOA on result patterns in visual stimuli.

Finally, we must acknowledge that the duration of each sequence varied as a function of the number of items within it. Thus, we cannot discard the interpretation that participants evaluated the overall duration of our visual and auditory sequences. However, co-variation between the numerosity and other physical characteristics is not specific to sequential presentations. Indeed, in the classic studies on enumeration (for a review, see Trick & Pylyshyn, 1994), the number of visual items in a simultaneous presentation is confounded with the amount of surface area covered by the stimuli. Similarly, in tactile perception, the overall intensity of the stimulation increases with the number of stimuli. Thus, for any type of stimuli, discontinuity in enumeration is observed when some extraneous cues are available. Future research on both sequential and simultaneous presentations has to address the question on the emergence of similar phenomena if these cues were controlled, as it is done in studies evaluating the numerical competence in infants (e.g., Xu & Spelke, 2000).

Acknowledgements

We thank Betty Bellanger for running the experiment and, two anonymous reviewers and Gerry Altmann for their comments on previous draft of this manuscript. Part of this work was done when the first author was invited fellow at University of Bristol funded by the Royal Society. The second author was funded by ANR.

References

Bregman, A. S. (1990). *Auditory scene analysis: The perceptual organization of sound*. Cambridge, MA: MIT Press.

Camos, V., Barrouillet, P., & Fayol, M. (2001). Does the coordination of verbal and motor information explain the development of counting in children? *Journal of Experimental Child Psychology, 78*(3), 240–262.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences, 24*, 87–185.

Cowan, N. (2005). *Working memory capacity*. Hove: Psychology Press.

Dehaene, S., & Cohen, L. (1994). Dissociable mechanisms of subitizing and counting: Neuropsychological evidence from simultanagnosic patients. *Journal of Experimental Psychology: Human Perception and Performance, 20*(5), 958–975.

Gagnon, L., Mottron, L., Bherer, L., & Joanette, Y. (2004). Quantification judgement in high functioning autism: Superior or different? *Journal of Autism and Developmental Disorders, 34*(6), 679–689.

Gallace, A., Tan, H. Z., & Spence, C. (2006). Numerosity judgments for tactile stimuli distributed over the body surface. *Perception, 35*(2), 247–266.

Garner, W. R. (1951). The accuracy of counting repeated short tones. *Journal of Experimental Psychology, 41*(5), 310–316.

Gelman, R., & Tucker, M. F. (1975). Further investigations of the young child's conception of number. *Child Development, 46*(1), 167–175.

Jevons, W. S. (1871). The power of numerical discrimination. *Nature, 3*, 281–282.

John, I. D. (1972). Some variables affecting judgments of auditory temporal numerosity. *Australian Journal of Psychology, 24*(3), 347–352.

Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *American Journal of Psychology, 62*, 498–525.

Lépine, R. (2003). Le processus de subitizing chez les enfants présentant un retard mental [Subitizing process in children with mental retardation]. *Approche Neuropsychologique des Apprentissages chez l'Enfant, 15*, 256–263.

Mandler, G., & Shebo, B. J. (1982). Subitizing: An analysis of its components processes. *Journal of Experimental Psychology: General, 111*(1), 1–22.

Massaro, D. W. (1975). Backward recognition masking. *Journal of the Acoustical Society of America, 58*(5), 1059–1065.

Massaro, D. W. (1976). Perceiving and counting sounds. *Journal of Experimental Psychology: Human Perception and Performance, 2*(3), 337–346.

Peterson, S. A., & Simon, T. J. (2000). Computational evidence for the subitizing phenomenon as an emergent property of the human cognitive architecture. *Cognitive Science, 24*(1), 93–122.

Repp, B. H. (2007). Perceiving the numerosity of rapidly occurring auditory events in metrical and nonmetrical contexts. *Perception & Psychophysics, 69*(4), 529–543.

Riggs, K. J., Ferrand, L., Lancelin, D., Fryziel, L., Dumur, G., & Simpson, A. (2006). Subitizing in tactile perception. *Psychological Science, 17*(4), 271–272.

Trick, L. M., & Pylyshyn, Z. W. (1994). Why small and large numbers enumerated differently? A limited-capacity preattentive stage in vision. *Psychological Review, 101*, 80–102.

Turvey, M. T. (1973). On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review, 80*(1), 1–52.

Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition, 74*, B1–B11.