

Is approximate numerical judgment truly modality-independent? Visual, auditory, and cross-modal comparisons

Midori Tokita · Yui Ashitani · Akira Ishiguchi

Published online: 3 August 2013
© Psychonomic Society, Inc. 2013

Abstract The numerosity of any set of discrete elements can be depicted by a genuinely abstract number representation, irrespective of whether they are presented in the visual or auditory modality. The accumulator model predicts that no cost should apply for comparing numerosities within and across modalities. However, in behavioral studies, some inconsistencies have been apparent in the performance of number comparisons among different modalities. In this study, we tested whether and how numerical comparisons of visual, auditory, and cross-modal presentations would differ under adequate control of stimulus presentation. We measured the Weber fractions and points of subjective equality of numerical discrimination in visual, auditory, and cross-modal conditions. The results demonstrated differences between the performances in visual and auditory conditions, such that numerical discrimination of an auditory sequence was more precise than that of a visual sequence. The performance of cross-modal trials lay between performance levels in the visual and auditory conditions. Moreover, the number of visual stimuli was overestimated as compared to that of auditory stimuli. Our findings imply that the process of approximate numerical representation is complex and involves multiple stages, including accumulation and decision processes.

Keywords Approximate numerosity · Sensory modality · Cross-modal comparison · Accumulator model

In the 1940s and 1950s, several studies investigated how signals in Morse code could be learned and what type of signal (tones or light flashes) was more effective for correctly receiving the code (e.g., Browne, 1951; Morsh & Stannard, 1947).

An important factor in these tasks was enumeration of the numbers of tones or flashes. Since then, “temporal numerosity” has been extensively investigated, and it has been concluded that the auditory modality is more effective than the visual modality for enumeration of successive signals (e.g., Lechelt, 1975).

Many previous studies have supported the idea that adults, as well as infants and animals, possess innate neural mechanisms that generate approximate numerical representations (e.g., Cantlon & Brannon, 2006; Feigenson, Dehaene, & Spelke, 2004; Hauser, Tsao, Garcia, & Spelke, 2003; Whalen, Gallistel, & Gelman, 1999). Furthermore, converging empirical findings from several areas of cognitive neuroscience have suggested that biologically determined mechanisms are responsible for approximate numerical representation (e.g., Nieder & Dehaene, 2009; Piazza, 2010). One of the claims made by the proponents of this theory is that the processing of approximate numerical representation is independent of sensory modality, and that it could genuinely represent the numerosity of any set of discrete elements, whether they were presented in a visual or auditory condition (e.g., Barth, Kanwisher, & Spelke, 2003; Gallistel & Gelman, 1992). These studies demonstrated that people could compare the numbers of stimuli across different modalities in the same way as within modalities. It was claimed that comparisons across presentation modalities are not performed using modality-specific numerical representations, but rather using a true abstract numerical representation system. In addition, evidence for modality-independent numerical ability has been claimed in infants (e.g., Jordan & Brannon, 2006; Jordan, MacLean, & Brannon, 2008; Kobayashi, Hiraki, & Hasegawa, 2005) and animals (Beran, 2012; Jordan, Brannon, Logothetis, & Ghazanfar, 2005). Although most of the studies exploring cross-modal abilities in infancy have involved small sets, many studies investigating discrimination abilities involving exclusively auditory (e.g., Lipton & Spelke, 2003, 2004) and exclusively visual (e.g., Xu & Spelke, 2000) stimuli have revealed similar ratio signatures across modalities.

M. Tokita (✉) · Y. Ashitani · A. Ishiguchi
Graduate School of Humanities and Sciences, Ochanomizu
University, 2-1-1 Otsuka, Bunkyo-ku, Tokyo 112-8610, Japan
e-mail: tokita.midori@ocha.ac.jp

The modality independence of numerical judgment has been explained by the *accumulator*, or *mode-control*, model. This model accounts for the behavioral characteristics of numerical estimation or discrimination of sequentially presented stimuli in humans and animals (Brannon & Roitman, 2003; Cordes, Gelman, Gallistel, & Whalen, 2001; Nieder & Dehaene, 2009). Figure 1 shows a schematic view of the accumulator model for the numerosity discrimination task. This model comprises a nonverbal counter (accumulator), memory, and a comparator. At the onset of a stimulus, pulses are gated into an accumulator (i.e., a nonverbal counting process), which integrates the number of counts over time, and the accumulated value is then transferred to memory. Humans and animals compare the current value in the accumulator with the stored value in order to determine what response to make (i.e., a decision process). Cordes et al. (2001) described the accumulation of numerical magnitude using the metaphor of placing “one cup for each item or event” into the accumulator. It has been claimed that in this model, accumulator noise accounts for the scalar variability, or more generally, the Weber law characteristic in numerical estimation and discrimination.

It has remained unclear, however, whether these approximate numerical representations are truly modality-independent. We have three primary reasons for doubting the modality independence of approximate numerical representations. First, some evidence has shown significant differences in the performance of numerical judgments using the visual, auditory, and tactile senses (e.g., Lechelt, 1975; Philippi, van Erp, & Werkhoven, 2008; Riggs, Ferrand, Lancelin, Fryziel, & Dumur, 2006). For example, in a rapid counting experiment, Lechelt compared the adult performance of numerosity judgments of visual, auditory, and tactile stimuli and demonstrated that perceived numerosity differed among the modalities. Philippi et al. demonstrated that stimuli with a short interstimulus interval (ISI) were underestimated and that this tendency was stronger for the visual than for the auditory stimuli. Although the target of investigation in these studies was exact counting, not approximate judgments,

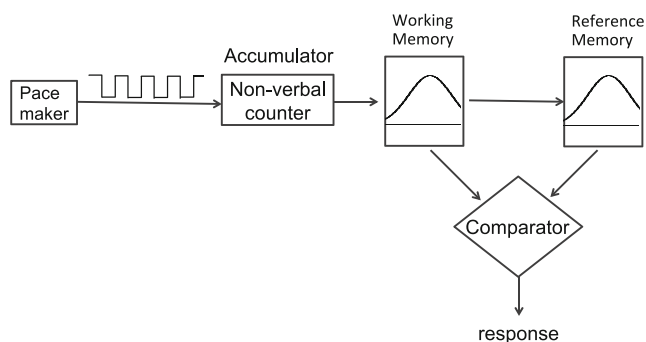


Fig. 1 Diagram of the accumulator model for numerosity discrimination of sequentially presented stimuli. The diagram was created with reference to Meck and Church (1983), Cordes et al. (2001), and Nieder and Dehaene (2009)

the authors implied that participants might shift their strategies from accurate counting to an estimation strategy when the presentation rate was faster. In infant studies, Mix, Levine, and Huttenlocher (1997) questioned the claims that infants can represent the numerosities of sets in different modalities, demonstrating that when the presentation rates and durations of auditory sequences were randomly varied, no significant preference for the equivalent or nonequivalent visual display was observed.

Second, it is known that the processing of temporal information is much more efficient in the auditory than in the visual modality (Ivry & Schlerf, 2008; Penney, Gibbon, & Meck, 2000). For example, in time-related tasks such as duration discrimination and empty-interval estimation, performance using auditory presentation was significantly better than that using visual and tactile presentations (e.g., Grondin, 2010). Furthermore, since temporal information affects numerical discrimination (Tokita & Ishiguchi, 2011), the performance of numerical discrimination may differ among modalities.

Third, limitations may exist within the experimental procedures of the empirical studies that have claimed modality independence for numerical representation, in terms of stimulus presentation, precision of measurements, and the number of stimuli tested. In relation to stimulus presentation, some of the experiments presented visual stimuli simultaneously, whereas auditory stimuli were presented sequentially (e.g., Jordan & Brannon, 2006; Jordan et al., 2008). In this experimental procedure the true modality effect could not be tested, because the presentation format was confounded with presentation modality. Both auditory and visual stimuli should be presented sequentially in order to exclusively assess the modality. In relation to measurement, the measurement precision level in previous research has not been sensitive enough to reveal differences in precision across modalities. For example, Barth et al. (2003) found that the accuracy of cross-modal comparison tasks was comparable with that of intramodal tasks, suggesting that nonnumerical cues did not play a substantial role, even in intramodal tasks. However, the numerical contrasts in their studies were significantly large, with a Weber fraction of .50 or greater. With this level of measurement precision, differences in the performance of each task could remain undetected. More to the point, the numbers of items tested in exclusively sequential presentations were smaller than three in infant studies and five in animal studies, respectively. Because it remains unclear whether the system for representing small numbers of objects is distinct from that for representing larger numbers of objects, it will be necessary to test whether the effects of sensory modality differ among a variety of numerosities.

In this study, we tested whether and how the numerical comparisons of visual, auditory, and cross-modal presentations would differ under adequate control of the factors discussed

above. We presented stimuli in both modalities in a sequential manner, in which the items in a set were presented one by one. More importantly, we applied rigid control of the stimuli so that other properties, such as stimulus duration and interval duration, would not be confounded with the number of stimuli. To assess the numerical judgment performance in each modality condition, we measured the Weber fractions to test the differences in precision and the point of subjective equality (PSEs) in order to test the accuracy of numerical discrimination. We also tested larger numbers of items, using standard numbers of 10 and 20.

In Experiment 1, we compared numerical discrimination performance between visual and auditory presentations. In Experiment 2, we compared performance across visual, auditory, and cross-modal numerical presentations in order to examine how numerical information in the different modalities may integrate.

Experiment 1

We examined the precision of approximate numerical comparisons in two sensory modalities: visual and auditory. A schematic view of the stimulus presentation is shown in Fig. 2. In the visual condition, the stimuli in a set consisted of a sequence of flashes, whereas in the auditory condition they consisted of a sequence of tones. To examine precision, we obtained Weber fractions that would indicate the variance in the participants' numerical comparisons. To examine accuracy, we obtained PSEs, which indicate the bias. In deriving the Weber fractions and PSEs, we used the method of constant stimuli, in which the participants in each trial decided which stimulus—the standard or the comparison—had more events. We employed two levels of standard number, 10 and 20, in

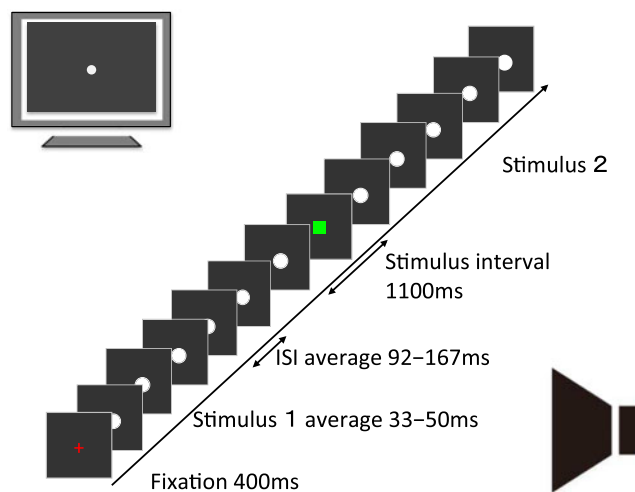


Fig. 2 Schematic view of the stimuli used in Experiments 1 and 2. Pairs of events were presented sequentially in a random order

order to test whether and how precision across presentation conditions would differ across standard numbers.

Method

Participants Nine participants took part, all of whom had normal or corrected-to-normal hearing and vision, and none of whom had any prior experience in numerical comparison tasks.

Design Two independent variables were examined: the sensory modality (i.e., visual and auditory) and standard number (i.e., 10 and 20). The comparison stimuli for the standard numbers of 10 and 20 were 8, 9, 11, and 12 and 16, 18, 22, and 24, respectively. Trials in the visual and auditory conditions were separated and constituted into trial blocks. The two experimental conditions were presented in a counterbalanced order. Trials in all standard number sets were intermixed within a block. Each condition had 320 trials (40 repetitions \times 4 comparison levels \times 2 standard numbers), resulting in 640 trials in total. Each block had 64 trials, with 10 blocks in total. The participants performed three or four blocks in each experimental session, for three sessions in total. Intermissions of approximately 3 min were given between blocks. The order of trials was completely randomized within a block. The standard stimuli came first in half of the trials and second in the remaining trials. The participants were given 16 practice trials before the actual experiment began.

Stimuli In the visual condition, two sequences of light gray dots appeared in a dark gray display region. The dots had a difference-of-Gaussians profile, with a bright center, a contrast of 0.5, and size was approximately 10 min of arc in diameter at the viewing distance of 115 cm. The luminances of the dot and of the background display region were 8 and 0.6 cd/m², respectively. In the auditory condition, two sequences of tones were presented using the built-in speaker of a desktop computer at an intensity of approximately 60 dB SPL. The auditory stimuli were 700-Hz pure sounds generated by a Macintosh computer. The sounds had a sample rate of 12 kHz and 16-bit resolution. The onset and offset of sounds were not amplitude modulated.

In both conditions, we carefully controlled the stimulus duration and the interstimulus interval (ISI) so that the duration of a sequence and the stimulus presentation rate would not be confounded with the number of stimuli. All stimuli in a particular sequence had the same duration, but the durations varied from sequence to sequence between 33 and 50 ms. In half of the trials in a block, the average ISI was 125 ms in both the standard and comparison sequences. In the remaining half, the average ISI in the comparison sequence was carefully controlled, so that the average total intervals for the standard and comparison sequences would be approximately equal.

Thus, the number of stimuli would be the only cue for numerical judgments.

In determining the ISI, two significant concerns were taken into consideration. First, the minimum ISI was set to 75 ms, so that the discriminabilities of individual elements in different modalities were equal. Many studies have provided evidence that the minimum ISI between two successive stimuli for correctly reporting their temporal order is about 40 ms, and that this threshold is invariant for auditory and visual stimuli (e.g., Kanabus, Szlag, Rojek & Pöppel, 2002; Pöppel, 1997). By setting a minimum ISI that was sufficiently higher than the temporal order threshold, the sets of elements in this experiment were perceived as being successive, independent of the sensory modality. Second, the maximum ISI was carefully determined so that the participants would not make judgments on the basis of verbal counting and/or temporal patterns. To make verbal counting impossible, the longest stimulus interval was set to be less than 250 ms, as previous studies had demonstrated that participants could not rely on verbal or subverbal counting within that duration (e.g., Piazza, Mechelli, Price, & Butterworth, 2006; Tokita & Ishiguchi, 2011). To make the sequence aperiodic, we randomly added temporal jitter (−24, −17, −8, 0, 8, 17, or 24 ms) to each ISI in all cases, so that the temporal rate would not constitute a rhythmic pattern.

Measurements The PSEs and Weber fractions were measured using the method of constant stimuli. First, the numbers of events for the comparison stimuli were plotted on the x -axis, and the proportion of “greater” responses for each comparison stimulus was plotted on the y -axis. The plotted data points constructed the psychometric function approximated by a cumulative Gaussian function, on which the difference threshold was obtained. This difference threshold was defined as the smallest amount of the element number change, for which a correct response rate of 75 % was achieved. Weber fractions were obtained by dividing the difference thresholds by the standard numbers. The PSEs were obtained as the values of the locations on the psychometric function at which the standard and comparative choice probabilities were equal to 50 %. In this experiment, we obtained the standardized PSE, dividing the difference between the PSE and the standard value by the number of the standard value. When deriving the psychometric functions, data points were transformed into z scores, and least squares regression functions were fitted to the data points for each condition. To test the fit of the psychometric functions, we applied Pearson correlations to test the fit of the least squares regression functions.

Procedure Participants sat in a darkened room at a distance of approximately 115 cm from the presentation screen. A keypad was placed directly in front of the participants, on which they responded by pressing the “1” or the “3” key. Each trial started with a red fixation cross for 400 ms, followed by the first

sequence. Pairs of sequences—standard and comparison—were shown in succession in a random order. The two sequences were separated by an interval of 1,100 ms. The participants’ task was to choose which sequence, the first or second, contained more stimuli. Feedback with a short beep sound was given when the participants made an incorrect choice. At the beginning of each session, the participants were explicitly instructed to attend to the number of stimuli presented and to discriminate on the basis of the numerosity that they felt, and not by verbal counting. They were also instructed to observe the center of the monitor in the auditory condition in the same manner as in the visual condition.

A Macintosh G4 computer was used to generate the display and sound, and to record the data. Stimuli were presented on a 17-in. color monitor at a refresh rate of 120 Hz (SONY Color Graphic Display Model GDM-F400).

Results

Figure 3 shows the average psychometric functions for each standard number. Figure 4 shows the mean Weber fraction in each condition. The fits of the data points to

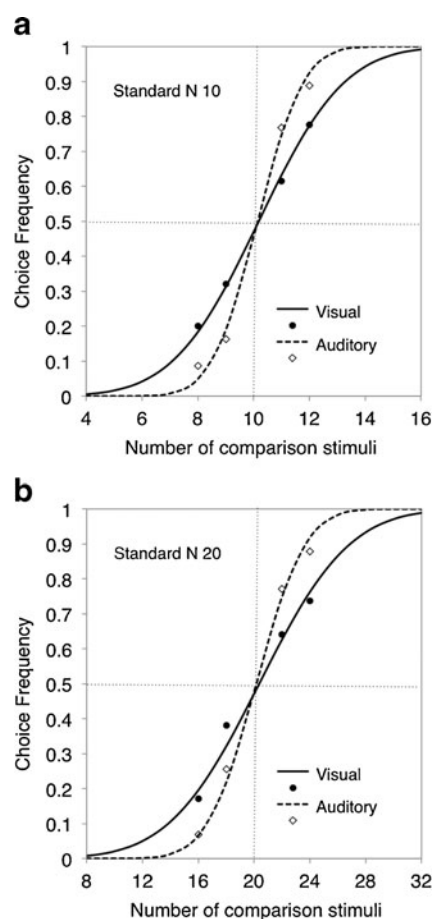


Fig. 3 Average psychometric functions for each presentation condition: **a** standard number of 10, **b** standard number of 20

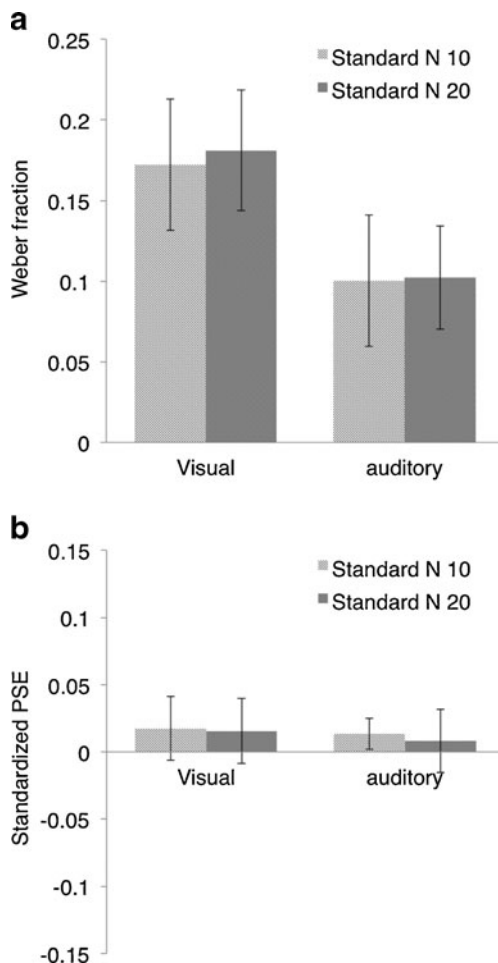


Fig. 4 Means of the **a** Weber fractions and **b** standardized points of subjective equality (PSEs) in the visual and auditory conditions at standard numbers of 10 and 20. Error bars represent standard deviations

the psychometric functions were generally good, and the Pearson product–moment correlation coefficient exceeded .9 in all cases.

To test whether and how the precision of numerical comparison differed between the visual and auditory conditions, a 2 modality (visual and auditory) \times 2 standard numbers (10 and 20) repeated measures analysis of variance (ANOVA) was conducted on the individual Weber fractions. This yielded a significant main effect of presentation modality, $F(1, 16) = 117.30$, $p < .0001$, $\eta^2 = .50$. Weber fractions in the auditory condition were significantly smaller than those in the visual condition, suggesting that numerical judgment was more precise in the auditory than in the visual modality. No significant effect of the standard numbers was observed, $F(1, 16) = 0.094$, $p = .76$, suggesting that the precision of numerical judgment was not affected by the number of stimuli within the numerical range tested in this experiment.

To test the accuracy of numerical comparison, we conducted a one-sample t test to compare the mean standardized PSEs of each condition with a PSE of 0. The mean PSE in the visual

condition was not significantly different from 0 at the standard numbers of 10, $t(8) = 2.08$, $p = .07$, and 20, $t(8) = 1.98$, $p = .08$. Similarly, the mean PSE in the auditory condition was not significantly different from 0 at the standard numbers of 10, $t(8) = 1.82$, $p = .11$, and 20, $t(8) = 0.81$, $p = .44$.

Discussion

What is the source of the difference in precision between the auditory and visual conditions? Two possibilities are considered here. First, counting errors (either skipping an item or counting it twice) might be greater in the visual than in the auditory condition, and greater counting errors could generate greater variability in numerical judgments. A second possibility was that there might be more accumulator noise in the visual than in the auditory condition. Greater accumulator noise could generate greater variability in numerical judgments. However, the counting error explanation is unlikely, because our results showed no effect of set size in either the auditory or the visual condition. When the number of stimuli that entered the counter increased, the number of steps would increase proportionally to the numerosity counted. If there were some probability of error at each count, the number of miscounts would increase with the number of steps. The variability in counts from this source should therefore obey binomial statistics. Using binomial statistics, the standard deviation of the numerical judgment can be formally expressed as

$$\sqrt{np(1-p)}.$$

Thus, the coefficient of variation (CV) of each stimulus condition is expressed as

$$\sqrt{np(1-p)/n}.$$

If $n_1 < n_2$, then

$$\sqrt{n_1p(1-p)/n_1} > \sqrt{n_2p(1-p)/n_2}.$$

Thus, the ratio of the standard deviation to the mean should be negatively correlated with the target number (Cordes et al., 2001). Therefore, if the source of the difference in variability were counting errors, the variability would be expected to decrease with the number of events to be enumerated, leading to a smaller Weber fraction with a standard number of 20 than with 10.

In Experiment 2, we tested how numerosity information would integrate in the cross-modal condition. If the difference in precision for the visual and auditory stimuli were due to a difference in accumulator noise, the performance of cross-

format comparisons would lie between the values for the visual and auditory conditions. This can be formally expressed as

$$\sigma_{AV} = \sqrt{\sigma_V^2 + \sigma_A^2}$$

where σ_V , σ_A , and σ_{AV} are the mean standard deviations of within-visual, within-auditory, and cross-modal comparisons. If $\sigma_V > \sigma_A$, as was the case in the results of Experiment 1, the relationship of the standard deviations among the three modality conditions would be predicted to be

$$\sigma_V > \sigma_{AV} > \sigma_A$$

Since the Weber fraction is obtained by dividing the standard deviation by the standard value, the Weber fractions of auditory and visual comparisons would be predicted to lie between those of the within-visual and within-auditory conditions. In Experiment 2, we tested this possibility by comparing the Weber fractions of within-visual, within-auditory, and between-modality numerical judgments.

Experiment 2

We tested the precision of approximate numerical comparisons in three presentation conditions (i.e., visual, auditory, and cross-modal). Since no systematic difference had been observed between the standard numbers in Experiment 1, we used one standard number (10) in this experiment. The stimulus presentations in the visual and auditory conditions were the same as those in Experiment 1. In the cross-modal condition, the stimuli in one set were presented as a visual sequence, and those in the other were presented as an auditory sequence. To examine precision, we obtained Weber fractions that indicated the variance in participants' numerical comparisons. To test the accuracy of the numerical comparisons, we derived the PSEs.

Method

Participants The experiment included nine newly recruited participants. None had any prior experience with numerical comparison tasks, and all had normal or corrected-to-normal hearing and vision.

Design We compared three presentation conditions: visual, auditory, and cross-modal. The cross-modal condition had two subconditions: Cross-modal Condition 1 and Cross-modal Condition 2. In Cross-modal Condition 1, the standard stimuli were visual sequences and the comparison stimuli were auditory sequences. In Cross-modal Condition 2, the standard stimuli were auditory sequences and the comparison stimuli were visual sequences. The numbers of comparison

elements for the standard number of 10 were 7, 8, 9, 11, 12, and 13. Trials in the cross-modal condition 1 and 2 were intermixed and constituted cross-modal trial blocks. Trials in the visual and auditory conditions were separated and constituted distinct trial blocks. The three experimental conditions were presented in a pseudocounterbalanced order. Each condition had 192 trials (32 repetitions \times 6 comparison levels), resulting in 768 trials in total. Each block consisted of 48 trials, with 16 blocks in total. Participants performed five to six blocks during each experimental session, for three session days in total. Intermissions of approximately 3 min were given between blocks. The participants were given 12 practice trials before the actual experiment began.

The stimuli, measurement, and procedures were the same as in Experiment 1, with the following exception: In both cross-modal conditions, the auditory and visual stimuli were shown in succession in a random order.

Results

Figure 5 shows the average psychometric functions for each condition. The fits of data points to the psychometric functions were generally good, and the Pearson product-moment correlation coefficient exceeded .9 in all cases. Figure 6 shows the means of the Weber fractions and of the standardized PSEs for all participants.

In order to test whether and how precision in numerical comparison differed among the visual, auditory, and cross-modal conditions, a four-condition repeated measures ANOVA was conducted on the individual Weber fractions. We found a significant main effect of presentation modality, $F(3, 8) = 14.86, p < .0001, \eta^2 = .51$, and a Bonferroni post hoc analysis revealed that the Weber fractions in the visual condition were significantly larger than those in the auditory condition ($p < .0001$), indicating that precision was substantially worse in the visual than in the auditory modality, as we had

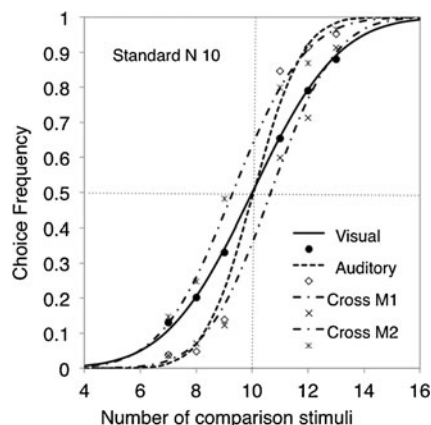


Fig. 5 Average psychometric functions for each presentation condition. “Cross M1” and “Cross M2” indicate Cross-modal Conditions 1 and 2, respectively

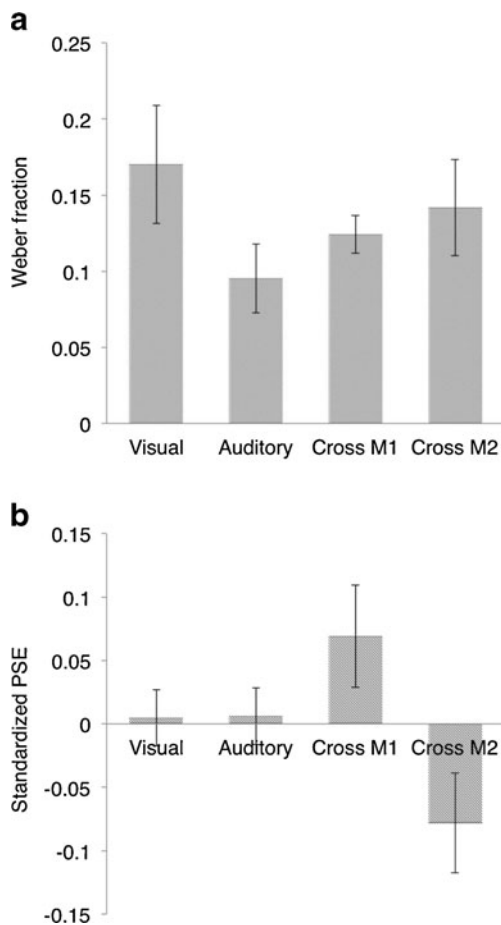


Fig. 6 Means of the **a** Weber fractions and **b** standardized points of subjective equality (PSEs) for each modality condition. “Cross M1” and “Cross M2” indicate Cross-modal Conditions 1 and 2, respectively. *Error bars* represent standard deviations

found in Experiment 1. The Weber fractions for Cross-modal Condition 2 ($p = .003$) were also larger than those in the auditory condition, and the Weber fractions tended to be larger in Cross-modal Condition 1 than in the auditory condition ($p = .12$). The Weber fraction was smaller for Cross-modal Condition 1 than for the visual condition ($p = .003$), and the Weber fraction for Cross-modal Condition 2 tended to be smaller than that for the visual condition ($p = .13$). No significance difference in Weber fractions was observed between Cross-modal Conditions 1 and 2 ($p = .82$). The results suggest that the performance of the cross-modal trials lay between the performance levels in the visual and auditory trials.

In order to test how cross-modal comparison affected the accuracy of numerosity discrimination, we conducted a one-sample t test to compare the mean standardized PSEs of each condition with a PSE of 0. The mean PSEs in the visual, $t(8) = 0.65$, $p = .53$, and auditory, $t(8) = 0.81$, $p = .44$, conditions were not significantly different from 0. The mean PSE in Cross-modal Condition 1 was significantly larger than 0, $t(8) = 5.12$, $p = .0009$, and the mean PSE in Cross-modal Condition 2 was

significantly smaller than 0, $t(8) = -5.93$, $p = .0003$. These results suggest that the number of visual stimuli was overestimated as compared to the auditory stimuli in both cross-modal conditions.

Discussion

The results clearly showed that the performance of cross-modal comparisons lay between the performance levels in auditory and visual trials, implying that the precision difference between the visual and auditory stimuli was due to a difference in accumulator noise.

The Weber fractions for the cross-modal conditions are consistent with the predictions of the accumulator noise hypothesis. However, the overestimation of visual stimuli was unexpected: A simulation with greater accumulator noise for visual numerosity predicted no bias in cross-modal comparison. What was the source of the overestimation? In the accumulator model, two error sources might cause the bias toward a greater response to visual stimuli: counting errors and decision errors. If the visual stimuli were counted twice with some probability and auditory stimuli were counted correctly, then overestimation of the visual stimuli might occur. However, this source of errors was implausible, as it had been tested in Experiment 1. Thus, decision bias at the comparator was more probable: The observer might tend to judge visual stimuli as being more numerous than auditory stimuli at the decision stage.

General discussion

We investigated approximate numerical judgments comparing visual, auditory, and cross-modal presentations using a carefully controlled experimental setting. In Experiment 1, we tested whether and how precision in numerosity discrimination would differ between visual and auditory conditions across two standard numbers, 10 and 20. In Experiment 2, we tested the precision and accuracy of visual, auditory, and cross-modal comparisons at a standard number of 10. Our results demonstrated three significant findings. First, the precision of numerical comparison of auditory sequences is significantly higher than that of visual sequences. Second, the performance of cross-modal comparison lies between the performance levels of visual and auditory trials. Third, in cross-modal comparisons, the number of visual stimuli is overestimated relative to the number of auditory stimuli. From these results, it can be assumed that each stimulus is stored in an accumulator in which visual and auditory numerosity have different signal variabilities (i.e., accumulator noise); the variability of visual signals is larger than that of auditory signals, and at the same time, the variability of accumulated magnitude becomes larger with visual numerosity than with auditory

numerosity, resulting in a larger Weber fractions in the visual condition. This assumption is strengthened by the fact that the variability of cross-modal comparisons lies between the variabilities of within-visual and within-auditory numerical comparisons. In the comparator, observers compare the current value in the accumulator with the stored value in the accumulator in order to determine the response to be made.

Our findings clearly demonstrate that the performance of approximate numerical judgments of sequentially presented stimuli depends on sensory modality, disagreeing with the claim of modality independence. The discrepancy between this and previous studies could be attributed to differences in the control of stimulus presentation, the accuracy of performance measurements, and the ranges of tested numerical values. It could be presumed that our rigid control of stimulus presentation and the accuracy of performance measurements revealed the genuine characteristics of visual and auditory numerical judgments. Moreover, the numerical values tested might be an important factor when considering how the presentation modality relates to numerical judgment. In this study, the numbers of elements in the standard stimuli were 10 or 20. In influential studies that have claimed the modality independence of numerical judgment in infants, only numerical values up to three have been tested (e.g., Jordan & Brannon, 2006; Jordan et al., 2005). In future research, it will be necessary to investigate whether the performance in these studies was truly related to an abstract numerical representation system, or whether it rather related to other cognitive mechanisms, such as memory capacity and/or matching ability.

The results are somewhat consistent with those of previous studies that have shown a difference in counting precision across modalities, and lower precision for visual presentation (e.g., Lechelt, 1975; Philippi et al., 2008). Notably, similar effects were observed between an exact counting task and an approximate numerical judgment task.

An unexpected but intriguing finding from this study is the overestimation of visual stimuli in the cross-modal conditions. This finding is inconsistent with the results of counting studies by Lechelt (1975) and Philippi et al. (2008). In these studies, the number of visual stimuli was *underestimated* in comparison with those of auditory and tactile stimuli. However, Lechelt also found an interesting trend toward underestimation of visual stimuli across presentation rates: Visual counts increased at higher rates, after substantial underestimation at slower rates. To explain this phenomenon, he suspected that observers might use different number reckoning strategies, by shifting from exact counting to approximate estimation as the presentation rate increased.

What is the source of the difference in accumulator noise between stimuli presented in the visual and auditory modalities? Why did observers overestimate the numbers of visual relative to auditory stimuli? How does greater variability in the visual condition relate to the overestimations in this condition?

The common aspects of those numerical judgments are that they were time-related, irrespective of sensory modality. In time-related tasks, such as duration and empty-interval judgments, differences in precision and accuracy occur between visual and auditory stimuli (see Grondin, 2010, for a review): Auditory signals are perceived as being longer than visual signals of the same duration (e.g., Walker & Scott, 1981; Wearden, Todd, & Jones, 2006), and sensitivity to time is much higher (i.e., lower threshold, or less variability) when intervals are marked by auditory rather than visual signals (Grondin, 2003). It has been hypothesized that the rate of the pacemaker might differ between auditory and visual stimuli in such a way that the internal clock runs faster for auditory than for visual stimuli, and that this “clock speed” difference is the main source of differences in subjective duration across modalities (e.g., Penney et al., 2000; Ulrich, Nitschke, & Rammsayer, 2006). Incorporating these time-related modality differences, two possibilities could be considered for the theoretical interpretation of our results. One possibility is that the difference in the Weber fractions between the visual and auditory stimuli would be due to a difference in the variability of time perception across modalities. The temporal characteristics of the incoming signal may influence the certainty of the presence of the signal. Because variability is larger for visual than for auditory stimuli in time processing, it could be predicted that the noise for a visual signal would be larger than that for an auditory signal. Consequently, the accumulated value of the visual signals has greater variability than that of the auditory signals, resulting in larger Weber fractions. Overestimation of visual stimuli would occur at the decision stage. Since both the filled duration and the interval duration were perceived as being shorter for visual than for auditory stimuli, the visual sequence would appear shorter in duration, and therefore events would be perceived at a higher rate (for a given number of events) than for an auditory sequence; that is, people confound number with rate (or inverse time) when they make their decision about number, and thus overestimate the number of visual stimuli.

Another possibility is that the speed of the pulses into the accumulator might be faster for visual than for auditory stimuli. A higher speed of the pulses could cause an overestimation of visual stimuli, analogous to the speeding up of the clock causing a larger accumulator pulse in time perception. The greater accumulation of pulses into the accumulator would lead to both overestimation of visual number and increased variability in representations for the same value. The speeding up of the pulses might be attributable to greater attentional demands for the visual stimuli; because the temporal resolution for the visual stimuli is much lower than that for the auditory stimuli, observers may need to be more attentive to individuate the items in a visual sequence.

To test and elaborate our predictions, extensive investigations will be required. Although we have attempted to interpret our results primarily in relation to accumulator noise and

decision processes, we could not rule out the possibility of differences in memory across modalities. Another important factor is the effect of stimulus intensity in numerosity judgments. Many researchers have demonstrated that stimulus intensity affects perceived duration (e.g., Matthews, Stewart, & Wearden, 2011; Xuan, Zhang, He, & Chen, 2007). When we theorize about numerosity judgments in incorporating temporal processes, we need to consider how stimulus intensity would affect numerical magnitude and how differences in intensity across modalities might interact with numerical judgments.

In conclusion, this study has provided clear evidence that approximate numerical judgments differ between visual and auditory stimuli in adult humans. Although many studies support the idea that adult humans, infants, and nonhuman animals share a modality-independent numerical system, it remains unknown how numerical information from different modalities is stored in the accumulator and combined at the judgment stage. Our findings imply that the process of approximate numerical representation is complex and involves multiple stages, including accumulation and decision processes.

Author Note This research was partially supported by a Grant-in-Aid for Scientific Research (B) (23330214) provided by the Japan Society for the Promotion of Science to the third author. The authors are grateful for comments and suggestions on an earlier version of the manuscript by Massimo Grassi, Will Matthews, and Sara Cordes.

References

- Barth, H., Kanwisher, N., & Spelke, E. (2003). The construction of large number representations in adults. *Cognition*, *86*, 201–221. doi:10.1016/S0010-0277(02)00178-6
- Beran, M. J. (2012). Quantity judgments of auditory and visual stimuli by chimpanzees (*Pan troglodytes*). *Journal of Experimental Psychology: Animal Behavior Processes*, *38*, 23–29. doi:10.1037/a0024965
- Brannon, E. M., & Roitman, J. D. (2003). Nonverbal representations of time and number in animals and human infants. In W. H. Meck (Ed.), *Functional and neural mechanisms of interval timing* (pp. 143–182). Boca Raton, FL: CRC Press.
- Browne, R. C. (1951). The identification of Morse signals when heard against a continuous tone. *British Journal of Psychology*, *42*, 279–287.
- Cantlon, J. F., & Brannon, E. M. (2006). Shared system for ordering small and large numbers in monkeys and humans. *Psychological Science*, *17*, 401–406.
- Cordes, S., Gelman, R., Gallistel, C. R., & Whalen, J. (2001). Variability signatures distinguish verbal from nonverbal counting for both large and small numbers. *Psychonomic Bulletin & Review*, *8*, 698–707. doi:10.3758/BF03196206
- Feigenson, L., Dehaene, S., & Spelke, E. (2004). Core systems of number. *Trends in Cognitive Sciences*, *8*, 307–314. doi:10.1016/j.tics.2004.05.002
- Gallistel, C. R., & Gelman, R. (1992). Preverbal and verbal counting and computation. *Cognition*, *44*, 43–74. doi:10.1016/0010-0277(92)90050-R
- Grondin, S. (2003). Sensory modalities and temporal processing. In H. Helfrich (Ed.), *Time and mind II* (pp. 61–77). Göttingen, Germany: Hogrefe & Huber.
- Grondin, S. (2010). Timing and time perception: A review of recent behavioral and neuroscience findings and theoretical directions. *Attention, Perception, & Psychophysics*, *72*, 561–582. doi:10.3758/APP.72.3.561
- Hauser, M. D., Tsao, F., Garcia, P., & Spelke, E. S. (2003). Evolutionary foundations of number: Spontaneous representation of numerical magnitudes by cotton-top tamarins. *Proceedings of the Royal Society B*, *270*, 1441–1446. doi:10.1098/rspb.2003.2414
- Ivry, R. B., & Schlerf, J. E. (2008). Dedicated and intrinsic models of time perception. *Trends in Cognitive Sciences*, *8*, 273–280.
- Jordan, K. E., & Brannon, E. M. (2006). The multisensory representation of number in infancy. *Proceedings of the National Academy of Sciences*, *103*, 3486–3489.
- Jordan, K. E., Brannon, E. M., Logothetis, N. K., & Ghazanfar, A. A. (2005). Monkeys match the number of voices they hear to the number of faces they see. *Current Biology*, *15*, 1034–1038.
- Jordan, K. E., MacLean, E. L., & Brannon, E. M. (2008). Monkeys match and tally quantities across senses. *Cognition*, *108*, 617–625.
- Kanabus, M., Szélag, E., Rojek, E., & Pöppel, E. (2002). Temporal order judgment for auditory and visual stimuli. *Acta Neurobiologiae Experimentalis*, *62*, 263–270.
- Kobayashi, T., Hiraki, K., & Hasegawa, T. (2005). Auditory–visual intermodal matching of small numerosities in 6-month-old infants. *Developmental Science*, *8*, 409–419.
- Lechelt, E. C. (1975). Temporal numerosity discrimination: Intermodal comparisons revisited. *British Journal of Psychology*, *66*, 101–108.
- Lipton, J. S., & Spelke, E. S. (2003). Origins of number sense. Large-number discrimination in human infants. *Psychological Science*, *14*, 396–401.
- Lipton, J. S., & Spelke, E. S. (2004). Discrimination of large and small numerosities by human infants. *Infancy*, *5*, 271–290.
- Matthews, W. J., Stewart, N., & Wearden, J. H. (2011). Stimulus intensity and the perception of duration. *Journal of Experimental Psychology: Human Perception and Performance*, *37*, 303–313. doi:10.1037/a0019961
- Meck, W. H., & Church, R. M. (1983). A mode control model of counting and timing processes. *Journal of Experimental Psychology: Animal Behavior Processes*, *9*, 320–334. doi:10.1037/0097-7403.9.3.320
- Mix, K. S., Levine, S. C., & Huttenlocher, J. (1997). Numerical abstraction in infants: Another look. *Developmental Psychology*, *33*, 423–428.
- Morsh, J. E., & Stannard, A. F. B. (1947). Studies in international Morse code: III. Retention of training by Morse code operators. *Canadian Journal of Psychology*, *1*, 141–144.
- Nieder, A., & Dehaene, S. (2009). Representation of number in the brain. *Annual Review of Neuroscience*, *32*, 185–208. doi:10.1146/annurev.neuro.051508.135550
- Penney, T. B., Gibbon, J., & Meck, W. H. (2000). Differential effects of auditory and visual signals on clock speed and temporal memory. *Journal of Experimental Psychology: Human Perception and Performance*, *26*, 1770–1787.
- Philippi, T. G., van Erp, J. B. F., & Werkhoven, P. J. (2008). Multisensory temporal numerosity judgment. *Brain Research*, *1242*, 116–125.
- Piazza, M. (2010). Neurocognitive start-up tools for symbolic number representations. *Trends in Cognitive Sciences*, *14*, 542–551. doi:10.1016/j.tics.2010.09.008
- Piazza, M., Mechelli, A., Price, C. J., & Butterworth, B. (2006). Exact and approximate judgements of visual and auditory numerosity: An fMRI study. *Brain Research*, *1106*, 177–188. doi:10.1016/j.brainres.2006.05.104
- Pöppel, E. (1997). A hierarchical model of temporal perception. *Trends in Cognitive Sciences*, *1*, 56–61.
- Riggs, J. K., Ferrand, L., Lancelin, D., Fryziel, L., & Dumur, G. (2006). Subitizing in tactile perception. *Psychological Science*, *17*, 271–272.
- Tokita, M., & Ishiguchi, A. (2011). Temporal information affects the performance of numerical discrimination: Behavioral evidence for

- a shared system for numerical and temporal processing. *Psychonomic Bulletin & Review*, 18, 550–556.
- Ulrich, R., Nitschke, J., & Rammsayer, T. (2006). Crossmodal temporal discrimination: Assessing the predictions of a general pacemaker-counter model. *Perception & Psychophysics*, 68, 1140–1152.
- Walker, J. T., & Scott, J. (1981). Auditory–visual conflicts in the perceived duration of lights, tones, and gaps. *Journal of Experimental Psychology. Human Perception and Performance*, 7, 1327–1339. doi:10.1037/0096-1523.7.6.1327
- Wearden, J. H., Todd, N. P., & Jones, L. A. (2006). When do auditory/visual differences in duration judgements occur? *Quarterly Journal of Experimental Psychology*, 59, 1709–1724. doi:10.1080/17470210500314729
- Whalen, J., Gallistel, C. R., & Gelman, R. (1999). Nonverbal counting in humans: The psychophysics of number representation. *Psychological Science*, 10, 130–137. doi:10.1111/1467-9280.00120
- Xu, F., & Spelke, E. S. (2000). Large number discrimination in 6-month-old infants. *Cognition*, 74, B1–B11.
- Xuan, B., Zhang, D., He, S., & Chen, X. (2007). Larger stimuli are judged to last longer. *Journal of Vision*, 7(10):2, 1–5. doi:10.1167/7.10.2

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.