

Working memory resources are shared across sensory modalities

V. R. Salmela · M. Moisala · K. Alho

© The Psychonomic Society, Inc. 2014

Abstract A common assumption in the working memory literature is that the visual and auditory modalities have separate and independent memory stores. Recent evidence on visual working memory has suggested that resources are shared between representations, and that the precision of representations sets the limit for memory performance. We tested whether memory resources are also shared across sensory modalities. Memory precision for two visual (spatial frequency and orientation) and two auditory (pitch and tone duration) features was measured separately for each feature and for all possible feature combinations. Thus, only the memory load was varied, from one to four features, while keeping the stimuli similar. In Experiment 1, two gratings and two tones—both containing two varying features—were presented simultaneously. In Experiment 2, two gratings and two tones—each containing only one varying feature—were presented sequentially. The memory precision (delayed discrimination threshold) for a single feature was close to the perceptual threshold. However, as the number of features to be remembered was increased, the discrimination thresholds increased more than twofold. Importantly, the decrease in memory precision did not depend on the modality of the other feature(s), or on whether the features were in the same or in separate objects. Hence, simultaneously storing one visual and one auditory feature had an effect on memory precision equal to those of simultaneously storing

two visual or two auditory features. The results show that working memory is limited by the precision of the stored representations, and that working memory can be described as a resource pool that is shared across modalities.

Keywords Working memory · Attention · Load · Precision · Psychophysics · Discrimination threshold · Audio–visual · Auditory features · Visual features · Cross-modal

The origin of memory limitations has been a major issue of debate in working memory research. The classical model of working memory suggests that both the visual and auditory modalities have separate and independent stores of limited capacity for short-term maintenance (Baddeley, 2010; Baddeley & Hitch, 1974). The two subsystems, as well as other aspects of working memory, have been studied using a concurrent-task procedure in which participants perform two tasks simultaneously during memory encoding—for example, a digit span task and a reasoning task (Baddeley & Hitch, 1974). In such settings, the recall of verbal materials is more disrupted by a phonological task than by a nonphonological visual task, and the recall of visual materials is disrupted more by a spatial task than by a verbal task (Brooks, 1968). Furthermore, when the participant is instructed to continuously shadow (i.e., repeat aloud) spoken letters during memory encoding, this verbal task disrupts delayed recall of simultaneously heard verbal materials more than delayed recall of simultaneously seen visual material (Kroll, Parks, Parkinson, Bieber, & Johnson, 1970). In addition, a spatial task involving letters disrupts spatial tracking but not performance in a verbal task (Baddeley, 1986), and visual memory is less disturbed than auditory memory while simultaneously performing a backward-counting task (Scarborough, 1972). These results suggest separate working memory subsystems for visuospatial and verbal/phonological materials. However, the subsystems

V. R. Salmela (✉) · M. Moisala · K. Alho
Institute of Behavioural Sciences, Division of Cognitive and
Neuropsychology, University of Helsinki, FI-00014 Helsinki,
Finland
e-mail: viljami.salmela@helsinki.fi

M. Moisala
Department of Teacher Education, Institute of Behavioural Sciences,
University of Helsinki, Helsinki, Finland

K. Alho
Helsinki Collegium for Advanced Studies, University of Helsinki,
Helsinki, Finland

are not totally independent, since both are controlled by central executive functions, which explain subtle interference effects. An alternative to separate auditory and visual memory stores is that the limitation arises from some general mechanism or process with a limited capacity, regardless of the sensory modality—for example, from a limited focus of attention (Cowan, 1997, 2011). According to this view, working memory is a part of long-term memory that is activated by the focus of attention (Cowan, 1997, 2011).

A common view suggests that working memory is object-based and that memory or attention capacity is limited to three or four objects (Cowan, 2001). Early studies on visual working memory suggested a purely object-based memory store (Luck & Vogel, 1997, 1998), but later studies have shown that resources compete within a feature dimension, but are separate for different feature dimensions (Wheeler & Treisman, 2002), and that increasing the number of features to be remembered in each object decreases both memory performance and precision (Fougnie, Asplund, & Marois, 2010; Oberauer & Eichenberger, 2013; Olson & Jiang, 2002). Working memory seems not to be purely feature-based, either, since the locations and types of features within objects also affect memory: Features are remembered best when they are in the same spatial region of an object (same-object advantage), worse when they are in different parts of an object, and worst when they belong to different objects (Fougnie, Cormiea, & Alvarez, 2013; Huang, 2010b; Olson & Jiang, 2002; Xu, 2002). Furthermore, increasing the number of objects to be remembered decreases memory performance and memory precision, but increasing the number of features within each object decreases only precision (Fougnie et al., 2010; Fougnie et al., 2013). Thus, according to these studies, working memory capacity is not solely limited by the number of objects or features, but by the number of objects and by the number, location, and type of features within the memorized objects.

Cumulative evidence from both visual (Anderson & Awh, 2012; Anderson, Vogel, & Awh, 2011; Bays, Catalao, & Husain, 2009; Bays & Husain, 2008; Huang, 2010a; Murray, Nobre, Astle, & Stokes, 2012; Salmela, Lähde, & Saarinen, 2012; Salmela, Mäkelä, & Saarinen, 2010; Salmela & Saarinen, 2013; Wilken & Ma, 2004; Zhang & Luck, 2008) and auditory (Kumar et al., 2013) working memory studies suggests that the memory limitations with a small number of items are due to the precision of stored representations; that is, only a few items can be remembered with high precision, but several items can be remembered with lower precision. The trade-off between memory capacity and the precision of representations has been found for several visual features—for example, spatial location (Bays & Husain, 2008), orientation (Anderson et al., 2011; Bays & Husain, 2008; Salmela & Saarinen, 2013), contour shape (Salmela et al., 2012; Salmela et al., 2010; Zhang & Luck, 2008), and color (Anderson & Awh, 2012; Bays et al., 2009; Huang, 2010a). The precision of auditory working memory has been much less studied, but a similar type of trade-off has been found for tone pitch (Kumar

et al., 2013). In vision, it seems that the trade-off between the number of stored items and the precision of representations is driven by stimulus information at the encoding stage, and cannot be volitionally varied according to task demands (Murray et al., 2012; Zhang & Luck, 2011), except when only a few items are memorized (Machizawa, Goh, & Driver, 2012). The precision of encoding can, however, vary trial by trial (Fougnie, Suchow, & Alvarez, 2012; van den Berg, Shin, Chou, George, & Ma, 2012). These studies suggest that working memory resources are shared across either the visual or the auditory representations held in memory. We asked whether the resources can also be shared across representations in different sensory modalities, and whether this sharing affects memory precision.

Resources can be flexibly shared across the visual and auditory modalities during simultaneous memory tasks (Morey, Cowan, Morey, & Rouder, 2011). Some studies have revealed clear capacity costs due to cross-modal sharing of resources (Morey & Cowan, 2004, 2005; Morey et al., 2011; Saults & Cowan, 2007; Vergauwe, Barrouillet, & Camos, 2010), whereas others have not (Cocchini, Logie, Della Sala, MacPherson, & Baddeley, 2002; Fougnie & Marois, 2011). The former studies support the general-resource model, and the latter studies support domain-specific resources. To our knowledge, the costs of a cross-modal task on the precision of memory representations have not been addressed previously, and it is not yet known whether resource allocation across modalities affects the precision of memory representations.

To test resource sharing across modalities, we measured memory precision for two visual and two auditory features. In Experiment 1, both visual features were varied in the same sine-wave gratings, both auditory features were varied in the same sine-wave tones, and the gratings and tones were presented simultaneously (Fig. 1). In Experiment 2, two tones and two gratings were presented sequentially, and only one feature was varied in each object. This manipulation was used to control for the same-object advantage when comparing the intra- and cross-modal results from Experiment 1. In every condition, the stimulus—and hence the perceptual load—was identical, and only the participants' memory task was varied. The participants had to memorize one to four features.

Increasing the number of to-be-remembered features can have different effects on memory precision. If working memory contains separate object-specific representations in the visual and auditory modalities, we should find a larger decrease in precision in Experiment 2 than in Experiment 1, due to the increased number of objects. However, if working memory is based on feature representations and domain-general resources, we should find clear decreases in memory precision in both experiments, immediately as the number of features to be remembered increases.

Previous cross-modal memory studies have tested whether or not sharing resources between modalities decreases memory performance. A decrease in memory task performance has been

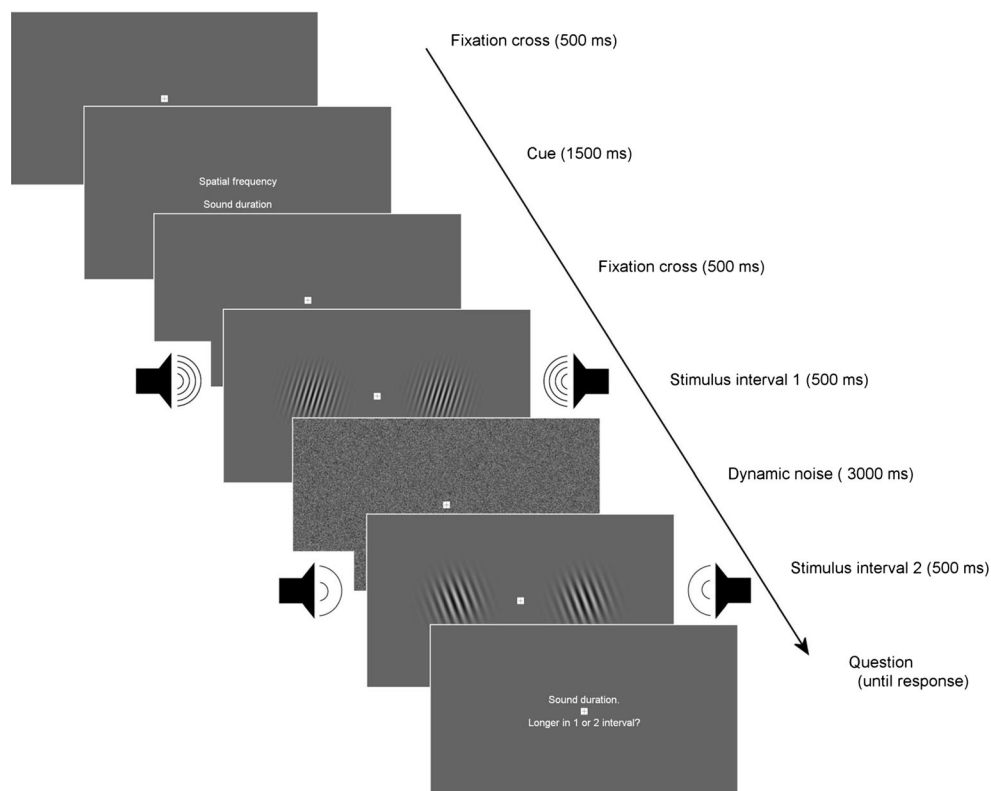


Fig. 1 Setup of Experiment 1. A two-interval forced choice setup was used to measure delayed discrimination thresholds for the spatial frequency and orientation of gratings and for the pitch and duration of simultaneous tones. The thresholds for the four features were measured separately, simultaneously, and in all possible two- and three-feature

combinations. In every condition, the stimuli varied in all four dimensions. A written cue specifying the one to four target features to be remembered was presented only at the beginning of the experiment and between blocks of 36 trials, except in the precue condition with four features, in which one random cue was presented on every trial

interpreted as support for domain-general resources, and the absence of a decrease in task performance has been interpreted as support for domain-specific resources. If memory resources are indeed domain-specific, then more resources would be available in cross-modal tasks, which in turn should lead to better memory performance in cross-modal than in intramodal tasks. If, on the contrary, memory resources are shared by modalities, then the intramodal and cross-modal conditions should not differ. Hence, in order to test the specificity of memory resources, the question should be turned around: Is there a cross-modal *benefit* in working memory tasks? We tested this by comparing the memory precision in intramodal and cross-modal conditions.

Method

Participants

Five participants with normal or corrected-to-normal vision and without any known hearing deficits participated in the experiments. The first author was one of the participants; the other participants were naive as to the purpose of the study. The participants gave written informed consent, and the

experiment was conducted according to the ethical standards of the Declaration of Helsinki and approved by the Ethics Committee of Helsinki and Uusimaa Hospital District.

Equipment and stimuli

The stimuli were sine-wave gratings and tones, as well as visual and auditory white noise. The visual stimuli were presented on a linearized LCD monitor (resolution $1,080 \times 1,920$ pixels). The size of the display was 28.9×48.8 deg at the viewing distance of 56 cm. Head position was held constant with a chinrest. The auditory stimuli were presented via closed headphones (Bayerdynamic 770).

A pair of sine-wave gratings in a Gaussian envelope were presented on the left and right sides of a fixation cross (black cross on a white background) at 5.4 deg eccentricity. The space constant of the Gaussian was 1.22 deg, and the size of the stimulus was 4.9 deg. The Michelson contrast of the grating was 0.9, and its phase was random. The orientation and spatial frequency of the grating varied between -45 and 45 deg and between 0.5 and 2.0 c/deg, respectively.

Sine-wave tones were presented to both ears. The intensity of the tones was 62 dB SPL, and the tones had 10-ms linear

fade-in and fade-out times. The frequency and the duration of the tones varied between 1000 and 1200 Hz and between 200 and 500 ms, respectively.

All of the stimuli were presented, and the sine-wave gratings and tones were generated, with Presentation software (Version 14.0, www.neurobs.com). White noise was generated with MATLAB (MathWorks Inc). The size of the dynamic visual white noise was 14.5×24.4 deg, and the root mean square contrast was 0.2 (standard deviation of the luminance divided with mean luminance). The white noise sound intensity was 33.5 dB SPL.

Procedure—Experiment 1

The precision of working memory representations was quantified as the discrimination threshold. Discrimination thresholds were measured using a two-interval forced choice task with an adaptive 2–1 staircase method yielding a 70.7 % threshold. In Experiment 1, each trial began with a fixation cross for 500 ms (see Fig. 1). Then sine-wave gratings and tones were presented with simultaneous onsets. The duration of the grating was always 500 ms, but the durations of the tones varied. During the 3,000-ms memory period, dynamic white noise (30×100 -ms frames) was presented on the screen. Throughout the measurement, white noise was also delivered to the headphones. After the memory period, a second stimulus interval was presented (Fig. 1). The spatial frequency and orientation of the visual stimulus and the pitch and duration of the auditory stimulus in the second interval were always either increased or decreased, relative to the first interval. The amount of change was either random or determined by the participant's responses. After the second stimulus interval, the participants were asked to compare one feature of the second stimulus with that of the first stimulus. The participant's task was to say whether (1) the grating in the first interval or the grating in the second interval had a higher spatial frequency; (2) the tone in the first interval or the tone in the second interval had a higher pitch; (3) the grating was rotated to the left or the right between the intervals (i.e., whether the orientation was more clockwise in the first or in the second interval); or (4) the tone had a longer duration in the first or in the second interval. According to the participant's response, the difference in the discriminated stimulus feature between the intervals was changed. After an incorrect answer the difference was increased, and after two consecutive correct answers the difference was decreased. The amount of change was 0.037 c/deg, 4 Hz, 2.5 deg, or 10 ms, for spatial frequency, pitch, orientation, or tone duration, respectively. Until the second reversal, the change steps were increased to 2.5-fold. The final threshold estimate was calculated as an average of the last four reversal points. Each threshold was measured with 36 trials, and the average number of reversal points was ten (varied between four and 16).

The thresholds were measured for each feature separately, for four features simultaneously, and for each combination of two and three features. The participants were instructed to memorize one to four features, but only one feature was tested on each trial. The tested feature was selected randomly. The conditions with two, three, and four features to be memorized were conducted in blocks of 36 trials, and the participant was allowed a break between the blocks. Hence, the total duration of the different conditions varied from 36 (one feature) to 144 trials (four features). The one- and four-feature conditions were repeated three times. In the beginning of the condition, and in the beginning of each block, a written cue was shown to the participant to remind him or her of the feature or features to be memorized. The feature that was tested on a given trial was indicated by a written cue during the response trial after the second stimulus interval (Fig. 1).

For a baseline discrimination threshold with one feature to be remembered, each participant's performance was measured for each feature with a short, 200-ms memory period. In the additional precue condition, memory for four features was measured, but a cue was presented in the beginning of each trial. This condition was similar to the one-feature condition, except that the feature to be remembered was switched on every trial. In total, the performance of each participant was measured in 60 memory blocks (2,160 trials) and 12 baseline blocks (432 trials). The measurements were done in several 1- to 2-h sessions on different days. The whole experiment took 6–8 h for each participant.

Procedure—Experiment 2

Experiment 2 was identical to Experiment 1, except that in Experiment 2 four separate objects—each containing only one varied feature—were presented sequentially. The visual objects were (1) a grating with fixed orientation (vertical) and variable spatial frequency, presented on the left side of the fixation, and (2) a grating with fixed spatial frequency (1.25 c/deg) and variable orientation, presented on the right side of the fixation. The auditory objects were (1) a sine-wave tone of 500 ms with variable pitch, presented to the left ear, and (2) a sine-wave tone of 800 Hz with variable duration, presented to the right ear. The spatial locations of the gratings and tones (left/right) were randomized across the participants. In the first interval, the stimuli were presented sequentially with a 500-ms interstimulus interval. The order of the two tones and the two gratings was random, but they were always presented alternately—that is, tone–grating–tone–grating. A partial-response procedure was used, and the second interval contained only one object, which contained the feature that was to be tested on a given trial. The memory duration was always 3,000 ms, and hence the duration of the noise mask was varied from 1,500 to 3,000 ms, depending on the serial order of the feature in the first interval. Conditions with one, two, and four features to be memorized were measured. In total, each participant measured 28 memory blocks (1,008 trials)

in one 2-h session. For all participants, Experiment 1 was measured before Experiment 2.

Results

Memory precision in Experiment 1

On average, across all features and participants, the discrimination thresholds increased rapidly in Experiment 1 as the number of features to be remembered increased from one to two (Fig. 2a). The thresholds for two features to be remembered were twofold, in comparison to a single

feature to be remembered. When the number of features was increased further, the participants' performance showed a further decrease, and the threshold for storing four features was 2.5-fold, in comparison to the one-feature threshold (Fig. 2a). The results for each feature separately were virtually identical to the average results. Memory precision, measured separately for spatial frequency, pitch, orientation, and tone duration, first decreased rapidly and then reached an asymptote (Fig. 2b). Four separate repeated measures analyses of variance (ANOVAs) showed that the increases of threshold associated with increasing the number of features to be remembered were statistically significant for spatial

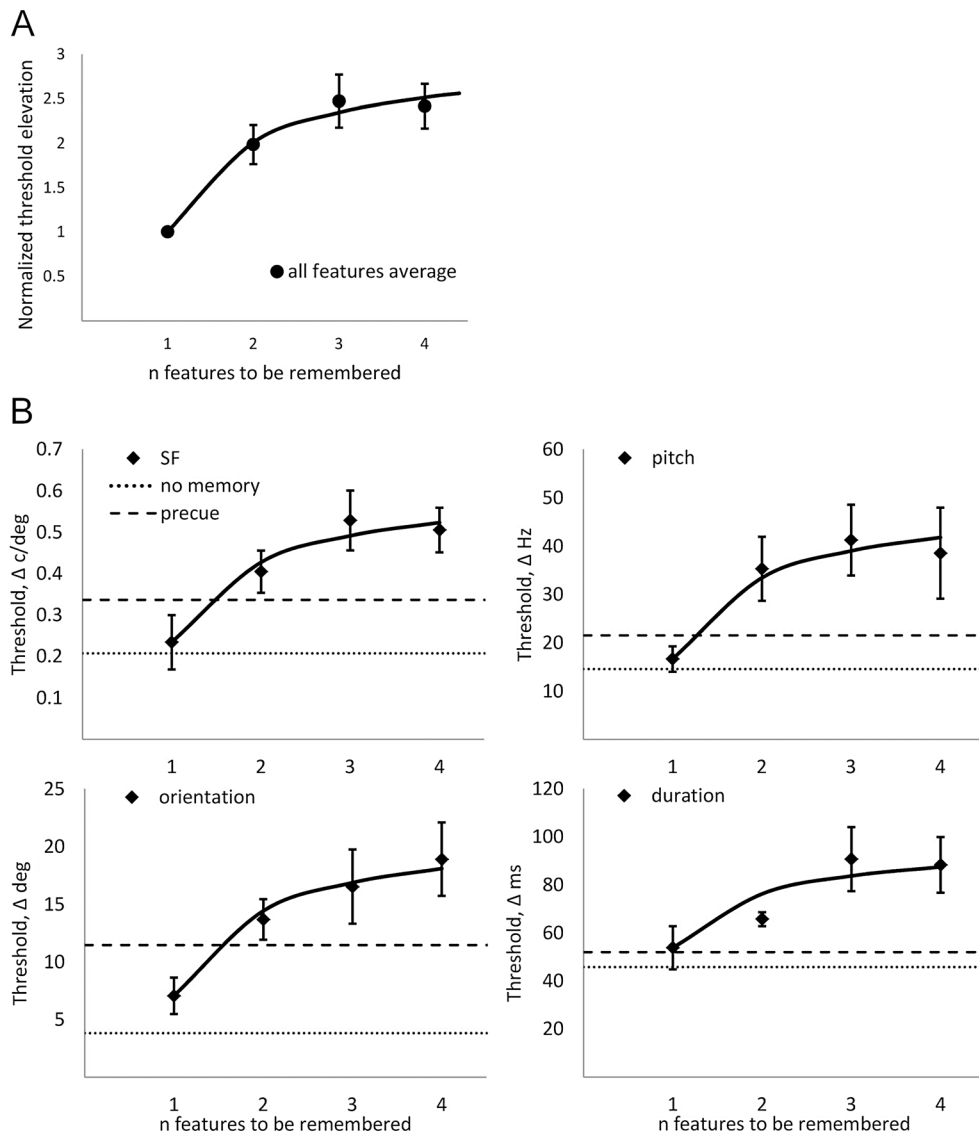


Fig. 2 Effect of memory load on precision in Experiment 1. (a) Average increase of thresholds as a function of memory load. The thresholds for four different features are normalized to one-feature conditions. The solid line depicts an asymptotic function fit to the data, and error bars depict standard errors of the means. (b) Results for the four features. The solid

lines depict an asymptotic function fit to the data, and error bars depict standard errors of the means. The dotted lines depict baseline without memory load, and the dashed lines depict a precue/attention-switching condition. SF=spatial frequency

frequency [$F(3, 12) = 13.886, p < .001$], pitch [$F(3, 12) = 10.244, p = .001$], orientation [$F(3, 12) = 5.95, p = .010$], and duration [$F(3, 12) = 7.012, p = .006$].

To quantify the effect of memory load on discrimination thresholds, an asymptotic function was fitted to the data:

$$Th = (1/n)Th_{n=1} + (1-[1/n])As, \quad (1)$$

where n is the number of stored features, $Th_{n=1}$ is the threshold with one item, and As is a constant corresponding to the asymptote of the function. The function predicts, on the basis of only the single-feature thresholds ($Th_{n=1}$) and resource sharing ($1/n$), the thresholds for two to four features. The asymptote (As) was the only free parameter, and was a scaling factor without any effect on the shape of the function. The function fit the data very well. The best fit for the average data (Fig. 2a) was obtained with an asymptote of 3.02 ($R^2 = .98$). For the separate features (Fig. 2b), the best fits were obtained for asymptotes of 0.62 c/deg, 50.19 Hz, 21.78 deg, and 98.70 ms, for spatial frequency, pitch, orientation, and tone duration, respectively.

The precision with a single feature was close to the perceptual precision (Fig. 2b, dotted lines) measured with the short (200-ms) retention interval. The single-feature memory threshold for orientation was slightly higher than the baseline threshold [$t(4) = 3.052, p = .038$], but for the other features the thresholds did not differ [$ts(4) = 0.739-0.778, ps > .48$ in all cases]. The baseline thresholds were also lower than the two-feature memory thresholds for spatial frequency, orientation, and pitch [$ts(4) = 4.007-5.503, ps < .016$ in all cases], but not for duration [$t(4) = 2.346, p = .079$].

Switching attention between features on every trial (Fig. 2b, dashed lines) had a relatively small effect on memory precision, since the thresholds in the precue condition were between the thresholds for one and two items. The thresholds in the precue condition were not higher than those in the single-feature memory condition [$ts(4) = -0.290$ to $1.995, ps > .117$ in all cases]. The precue thresholds were also not lower than the two-feature memory thresholds [$ts(4) = 1.118-1.353, ps > .248$ in all cases], except for pitch [$t(4) = 3.189, p = .033$]. In comparison to the baseline thresholds, the precue thresholds were higher than baseline for spatial frequency [$t(4) = 3.008, p = .040$] and orientation [$t(4) = 3.509, p = .025$], but not for pitch [$t(4) = 2.443, p = .071$] and duration [$t(4) = 0.722, p = .51$].

Effect of modality in Experiment 1

Increasing the number of features to be remembered increased the memory precision thresholds (Fig. 2). This suggests that memory resources are shared across the items currently stored

in memory. To test whether the resource sharing depends on the modalities of the features, the data were reanalyzed according to modality. If the memory resources are domain-specific, then there would be a cross-modal benefit in memory precision.

On average, across all features and participants, the discrimination thresholds were slightly higher in the cross-modal than in the intramodal conditions (Fig. 3a). Orientation discrimination thresholds were higher in the cross-modal condition than in the intramodal condition (Fig. 3b) when two features were to be remembered [$t(4) = 4.263, p = .013$], but not when three features were to be remembered [$t(4) = 2.059, p = .109$]. For spatial frequency and tone pitch and duration, the thresholds did not differ significantly [$ts(4) = -0.826$ to $0.595, ps > .445$ in all cases] in the intramodal and cross-modal conditions (Fig. 3b). Importantly, none of the conditions showed any benefit of cross-modality condition, which would be seen as lower thresholds in cross-modal than in intramodal conditions.

Memory precision in Experiment 2

In Experiment 2, four objects, each containing only one varied feature, were presented sequentially. The results were almost identical to those from the first experiment, in which stimuli were presented simultaneously. Again, the discrimination thresholds increased rapidly as the number of features to be remembered increased from one to two, and the thresholds reached an asymptote when four features were to be remembered (Fig. 4). The increases of the thresholds were statistically significant for spatial frequency [$F(2, 8) = 9.994, p = .007$], orientation [$F(2, 8) = 40.0552, p < .001$], and duration [$F(2, 8) = 16.848, p = .001$], but not for pitch [$F(2, 8) = 2.671, p = .129$]. Pitch discrimination thresholds showed only a shallow increase as a function of memory load. The best fit of the asymptotic function (Eq. 1) for the average data (Fig. 4a) was obtained with an asymptote of 3.24 ($R^2 = .99$). For each feature separately (Fig. 4b), the best fits were obtained for asymptotes of 0.54 c/deg, 33.67 Hz, 23.24 deg, and 103.11 ms, for spatial frequency, pitch, orientation, and tone duration, respectively.

Effect of modality in Experiment 2

On average, the discrimination thresholds were identical in the cross-modal and intramodal conditions of Experiment 2 (Fig. 5a). When analyzed separately for each feature, the discrimination thresholds for tone pitch were higher in the cross-modal than in the intramodal condition [$t(4) = 4.604, p = .010$; Fig. 5b]. The tone pitch threshold while memorizing two auditory features was identical to the threshold for tone pitch memorized alone, and the pitch thresholds were also identical when two or four cross-modal features were to be memorized

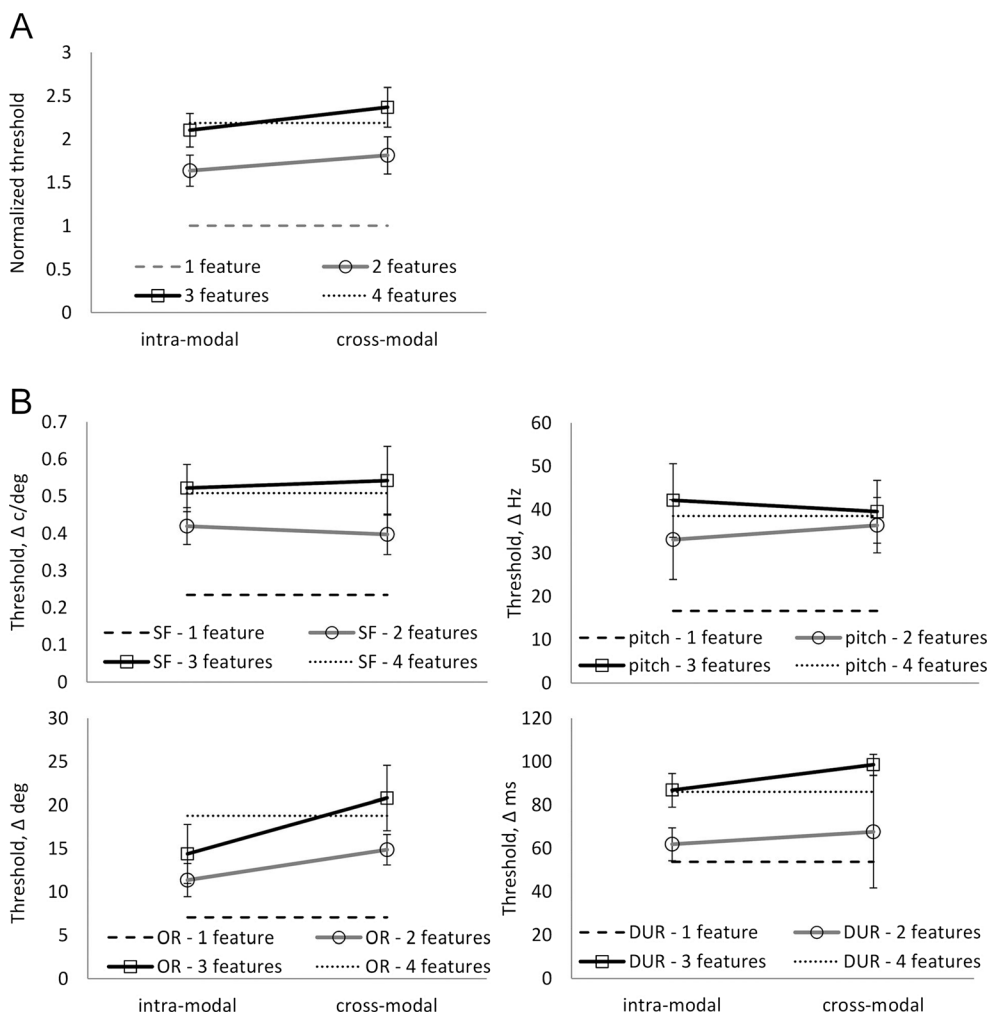


Fig. 3 Effect of sensory modality in Experiment 1. (a) Average change of threshold due to modality. The thresholds for four different features are normalized to conditions with one feature to be remembered. Error bars depict standard errors of the means. (b) Results for spatial frequency (SF), pitch, grating orientation (OR), and tone duration (DUR). In intramodal conditions, all or most attention was directed within a single modality, and in cross-modal conditions, attention was divided across modalities. One-feature conditions refer to trials in which a single feature was to be memorized. Intramodal two-feature conditions refer to trials in which the other feature was of the same modality (e.g., SF threshold measured together with OR threshold). Cross-modal two-feature conditions refer to

trials in which the other feature was of a different modality (e.g., SF threshold measured together with DUR threshold). Intramodal three feature conditions refer to trials in which two features shared a modality, and the third feature was from a different modality (e.g., SF and OR thresholds measured together with pitch threshold). Cross-modal three-feature conditions refer to trials in which the given feature had to be remembered with two features from a different modality (e.g., SF threshold measured with pitch and DUR thresholds), and four-feature conditions refer to trials in which all four features were to be memorized. The one- and four-feature data are actually intramodal and cross-modal points, but they are plotted as reference lines for visual clarity

(Fig. 5b). The orientation, spatial frequency, and tone duration discrimination thresholds did not differ significantly [$t_s(4) = -1.796$ to 0.243 , $p_s > .147$ in all cases] in intramodal and cross-modal conditions (Fig. 5b). Again, none of the conditions showed any cross-modal benefit.

Comparison of Experiments 1 and 2

To statistically test the difference between Experiments 1 and 2 (with simultaneous and sequential stimulus presentations, respectively), and the effects of different

numbers and types of features to be remembered, a three-way repeated measures ANOVA, with the factors Experiment, Number of Features, and Feature Type, was conducted for the normalized threshold values. The main effects of experiment [$F(1, 4) = 0.209$, $p = .671$] and feature type [$F(3, 12) = 2.569$, $p = .103$] were not statistically significant, but the effect of number of features was [$F(2, 8) = 15.980$, $p = .002$], as we had already observed in the previous analyses. The interactions were not significant [Experiment × Feature Type, $F(3, 12) = 1.213$, $p = .347$; Experiment × Number of

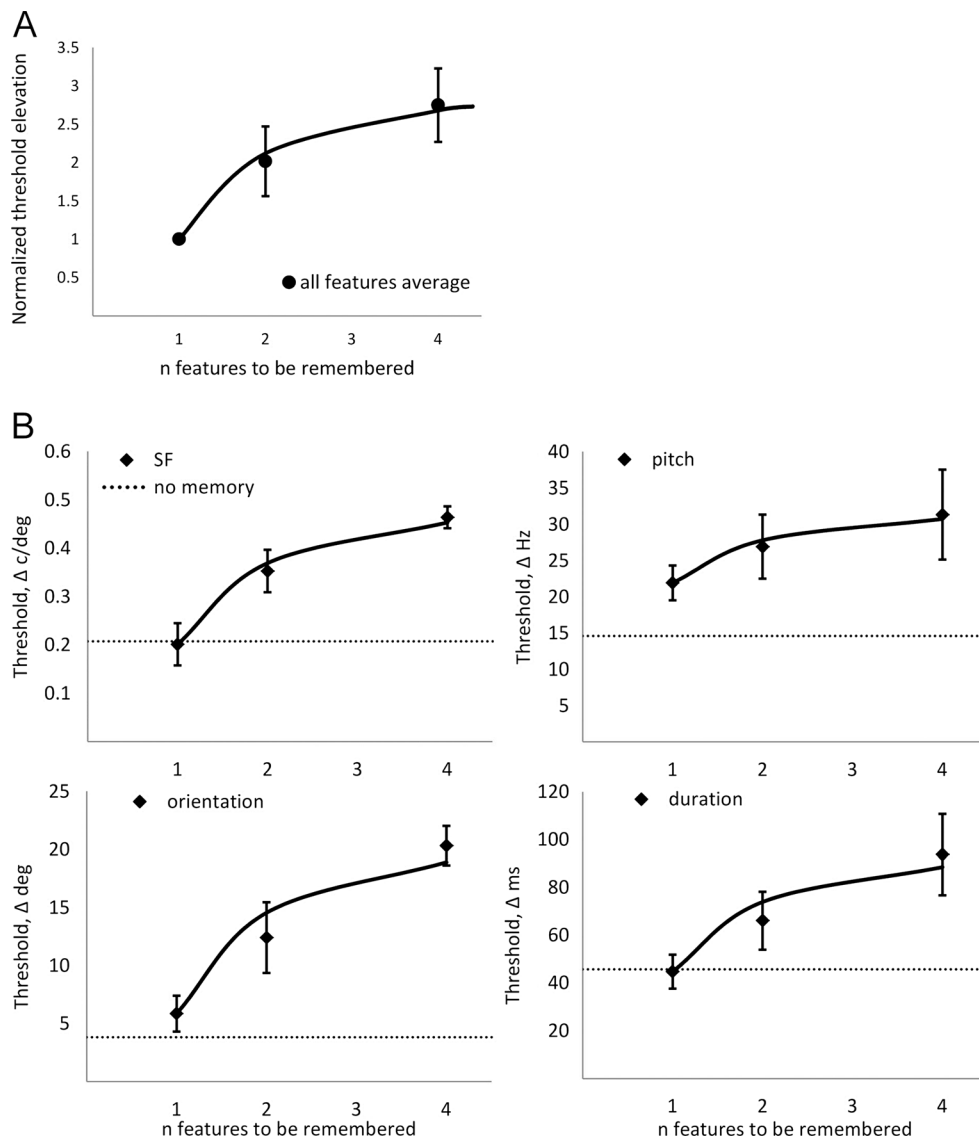


Fig. 4 Effect of memory load on precision in Experiment 2. (a) Average increase of thresholds as a function of memory load. The thresholds for four different features are normalized to one-feature conditions. The solid line depicts an asymptotic function fit to the data, and error bars depict

standard errors of the means. (b) Results for the four features. The solid lines depict an asymptotic function fit to the data, and error bars depict standard errors of the means. The dotted lines depict the baseline results without memory load measured in Experiment 1. SF=spatial frequency

Features, $F(2, 8) = 0.519, p = .614$; Experiment \times Feature Type \times Number of Features, $F(6, 24) = 1.273, p = .307$], except the Feature Type \times Number of Features interaction [$F(6, 24) = 3.166, p = .020$]. This significant interaction was due to the tone pitch thresholds, which significantly increased as a function of memory load in the first experiment, but not in the second experiment.

Two-way ANOVAs were conducted separately for each feature to test the effects of modality (two intramodal features vs. two cross-modal features) and experiment (simultaneous vs. sequential presentation). Only conditions in which two features were to be memorized were included in the analysis, because intramodality and cross-modality situations were clearly defined in those conditions. When the orientation of

the grating was to be remembered, the Modality \times Experiment interaction was statistically significant, but all other effects of cross-modality and experimental conditions were not (Table 1). The Modality \times Experiment interaction in orientation discrimination was due to the effect of modality found in Experiment 1, but not in Experiment 2.

Discussion

We tested whether working memory resources are shared across representations in the visual and auditory modalities. Our results showed that memory precision declined immediately as the number of features to be remembered increased to

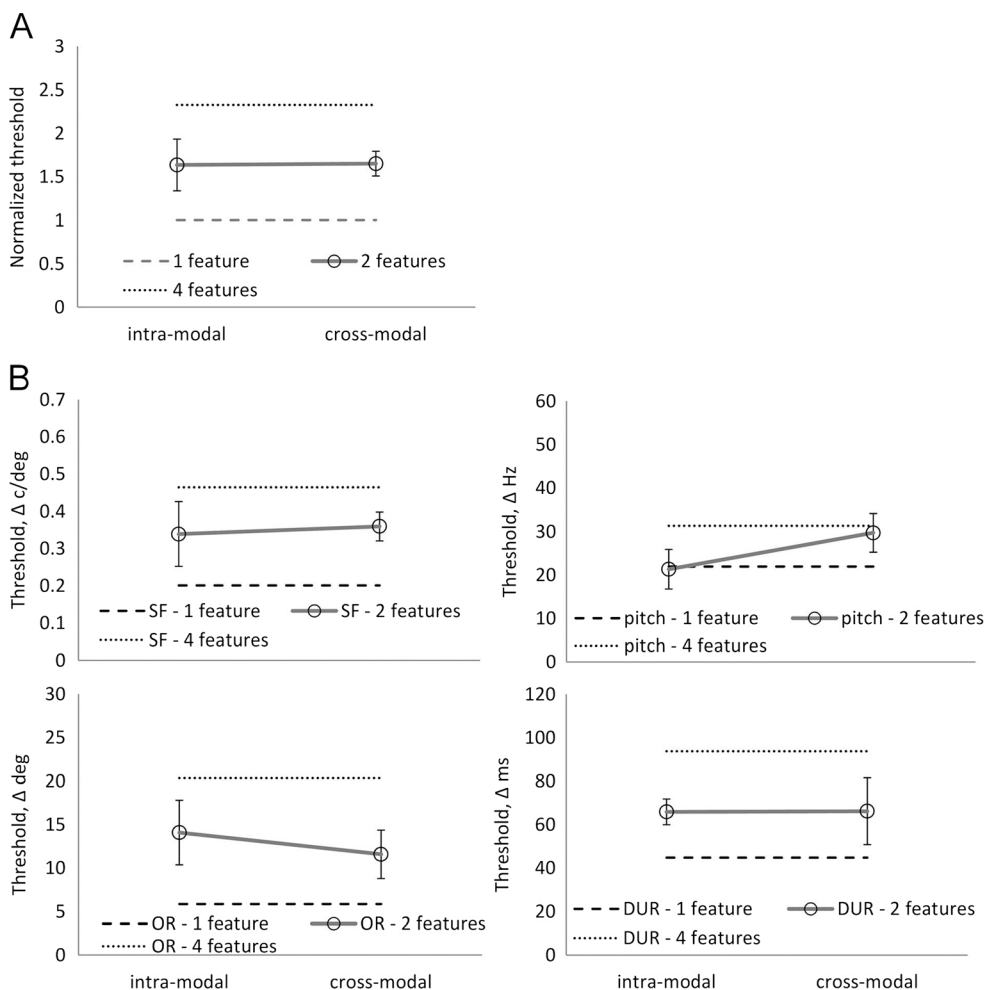


Fig. 5 Effect of sensory modality in Experiment 2. (a) Average change of threshold due to modality. The thresholds for four different features are normalized to conditions with one feature to be remembered. Error bars

depict standard errors of the means. (b) Results for spatial frequency (SF), pitch, grating orientation (OR), and tone duration (DUR). See the Fig. 3 caption for more details

more than one. The memory decline did not depend on the modality of the features (visual/auditory) or the feature type (spatial frequency, orientation, pitch, or duration) or on whether the features were in simultaneously or sequentially presented objects. Importantly, we observed no benefit of cross-modality conditions, and the precision was not better in cross-modal than in intramodal conditions.

Hence, the present results support a general-resource model of working memory instead of models containing modality-specific and object-based memory representations. The results suggest that working memory capacity is limited to only a few precise memory representations, and that the memory resources can be shared across sensory modalities with minimal costs in precision.

Table 1 Effects of cross-modality and experimental conditions

Memorized Feature	Modality (Intramodal vs. Cross-Modal)	Experiment (Simultaneous vs. Sequential)	Interaction (Modality × Experiment)
Spatial frequency	$F(1, 4) = 0.910, p = .394$	$F(1, 4) = 0.001, p = .983$	$F(1, 4) = 0.196, p = .681$
Pitch	$F(1, 4) = 3.608, p = .130$	$F(1, 4) = 2.444, p = .193$	$F(1, 4) = 0.513, p = .513$
Orientation	$F(1, 4) = 0.013, p = .915$	$F(1, 4) = 0.267, p = .633$	$F(1, 4) = 23.81, p = .008^*$
Duration	$F(1, 4) = 0.011, p = .920$	$F(1, 4) = 0.111, p = .756$	$F(1, 4) = 0.261, p = .636$

Two-way ANOVAs were conducted to test the effects of modality, experiment, and their interaction for each feature separately. Only conditions in which two features were to be memorized were included in the analysis. * $p < .05$

Working memory performance and capacity has mostly been studied by varying the number of items in the stimulus. However, in the present study, the stimulus was always the same—that is, it always contained a sine-wave grating and a sine-wave tone—and only the number of features to be memorized was varied. In agreement with several previous studies (Anderson & Awh, 2012; Anderson et al., 2011; Bays et al., 2009; Bays & Husain, 2008; Kumar et al., 2013; Salmela et al., 2012; Salmela et al., 2010; Salmela & Saarinen, 2013; Wilken & Ma, 2004; Zhang & Luck, 2008), the present results show that memory performance declines immediately as memory requirements increase. To our knowledge, the present study is the first to show that memory precision also declines in cross-modal conditions. For example, memory precision for visual orientation declined when the pitch of a tone was to be memorized simultaneously.

The present results are not compatible with working memory models that include object-based memory representations. According to object-based models, our memory task should have been trivial, since only a few objects were required to be held in memory. If working memory indeed has a capacity to store several objects, the task for two simultaneous objects should have been very easy, and there should not have been strong effects of increasing the number of items on memory precision. Instead, we found a clear and asymptotic increase in discrimination thresholds as a function of memory load. Previous studies have modeled the decrease of memory precision with different types of functions. The slots + averaging model of working memory suggests that when the number of items to be remembered is less than the number of slots, multiple samples of each item can be stored, and the memory precision is proportional to the square root of the number of samples (Palmer, 1990; Zhang & Luck, 2008). Likewise, resource-based models suggest a power-law relation between precision and the number of items to be remembered (Bays & Husain, 2008). Our asymptotic function is compatible with both of these accounts, but is slightly steeper in shape.

In the present study, we did not find a same-object advantage (Fougnie et al., 2013; Huang, 2010b; Olson & Jiang, 2002; Xu, 2002), and all of the features were remembered equally well when they were either within two object pairs presented simultaneously or within four objects presented sequentially. For tone pitch, in fact, memory precision was better in the sequential condition, showing a same-object *dis*-advantage. One difference between our experiments and previous studies using multiple features within visual objects was that we used narrow-band features. Our gratings were in a Gaussian envelope, and thus contained information only in narrow spatial-frequency and orientation bands. In contrast, previous studies have used sharp-edged stimuli (e.g., a square box with diagonal stripes) containing information across all spatial frequencies and multiple orientations. Perhaps the same-object advantage is only present when using broadband

stimuli. Another possibility is that the same-object advantage requires a larger number of features within each object. In our setup, we had only one or two features in each object to be memorized.

The present results are in contrast with the classical working memory model (Baddeley & Hitch, 1974) and with studies proposing independent, modality-specific memory subsystems. Baddeley's model contain visuospatial and phonological stores. However, instead of phonological stimuli, we used auditory tones, and it is therefore possible that the tones were not stored in the same storage as phonological material. Our experimental design was also quite different from previous studies using concurrent memory tasks that have found modality-specific effects (Baddeley, 1986; Brooks, 1968; Cocchini et al., 2002; Kroll et al., 1970; Scarborough, 1972). In those studies, the stimuli were typically word lists or numbers (therefore, involving semantic content), the tasks were continuous, and free recall was used to collect the responses. In our study, we used a very specific setup and low-level visual and auditory features, and directly tested whether resources could be shared across modalities. Another critical difference was that we used an adaptive method in our measurements. The adaptive method aims to keep task difficulty at the same level in all conditions and avoids possible ceiling and floor effects in the measurements. The decrease in memory precision as a function of memory load was the only general cost across all features that we found. This cost did not depend on whether features were in the same or in different modalities, or on whether the features were simultaneously or sequentially presented. In other words, the cost of adding a cross-modal feature to the memory task was identical to the cost of adding one intramodal feature to the memory task. Thus, we observed costs due to multitasking (i.e., to remembering multiple features simultaneously), but these costs did not depend on the modality.

In the present study, we wanted to test whether working memory resources are modality-specific or domain-general by using similar intramodal and cross-modal tasks. The key prediction of modality-specific resources is that in cross-modal conditions, performance should be improved, due to an increase in available resources. Contrary to this prediction, we did not find a clear cross-modal benefit in any condition, and thus our results suggest that the factor that limits the precision of working memory representations is domain-general. Our results are thus partly in agreement with the “focus of attention” model of working memory (Cowan, 2011). However, our results suggest that this domain-general resource is not object-based, but depends instead on the precision of representations. This suggests that the “size” of the attention focus is inversely related to the precision of representations. Previous studies using various combinations of multiple-object tracking, visuospatial, verbal, and semantic working memory tasks have found domain-general and

modality-specific, as well as attention-related, impairments in working memory performance (Fougnie & Marois, 2006; Morey & Cowan, 2004, 2005; Sauls & Cowan, 2007). Thus, it seems that with complex memory tasks, different types of interference can be found. With our simple adaptive-memory task, we found only domain-general effects. It might be argued that our results reflect the allocation of resources to modality-specific memory stores by executive operations. This is not likely to be the case, however, since our task was very simple—maintaining one bimodal object (or one to two auditory and one to two visual objects) in memory—and therefore demanded very little executive control. Furthermore, we measured the precision of memory representations, which presumably taps into the contents of memory stores.

During recent years, visual working memory has been studied with recall tasks in which the participant selects the feature value that most resembles the memorized item (e.g., green color) from a continuous set (e.g., a color wheel) of values (Wilken & Ma, 2004). From the distribution of participants' choices, the probability that participants stored the item (i.e., that the response was not random) and the precision of memory (i.e., the amount of error) can be quantified (Zhang & Luck, 2008). The latter value might describe memory precision more accurately, since random responses—for example, due to memory failures—have been removed. The division of responses into two different distributions (remembered items and random responses) is not always necessary, since random responses in a recall task may be absent up to at least six items (Salmela et al., 2012; Salmela & Saarinen, 2013). Nevertheless, the separation of these two values is most critical when the number of to-be-remembered items is large. In the present study, we used only a low number of features (one to four), and hence, our precision estimates are reliable.

It has been suggested that for audiovisual stimuli capacity limitations are fundamentally different, so that audiovisual integration is limited to only one object (Van der Burg, Awh, & Olivers, 2013). Although in our task integration was not required, a limited capacity of audiovisual processing might also have had an effect in our setup. It is possible that auditory and visual objects are automatically integrated when the objects are presented simultaneously (Degerman et al., 2007). However, our participants were able to perform the task well above chance level (since the adaptive method kept the performance at 70 % correct), even when four features were to be remembered. It has previously been shown that in a difficult visual memory task, only one item can be remembered precisely (Salmela et al., 2010). Instead of a fixed capacity for one object, these results suggest that when high precision is required in the task, all resources may be allocated to maintaining only a single representation.

There are no clear functional explanations for why working memory resources should be allocated flexibly across domains. One explanation is that the resources are dynamically distributed across all representations, to optimize the usage of limited capacity. An ideal-observer analysis based on optimal information transfer does indeed explain many properties of working memory (Sims, Jacobs, & Knill, 2012). Recently, it has been suggested that the limited capacity of attention and memory arise from cognitive “maps”—neural mechanisms analogous to retinotopic maps in the visual cortex (Franconeri, Alvarez, & Cavanagh, 2013). According to this model, the total capacity limit would be the size of the cognitive map, and flexible allocation of resources and interference between memory items would be inherent properties of these cognitive maps. Our results are compatible with this model. A common cognitive map containing representations from both visual and auditory modalities would explain the similarity of the present results for the intramodal and cross-modal conditions.

If working memory limitations are indeed due to cognitive maps, it still remains unclear in which way resources are shared across memory representations. The present and previous (Bays & Husain, 2008; Salmela et al., 2010; Salmela & Saarinen, 2013) results suggest that memory resources are evenly allocated to each item. This type of optimal resource sharing could also be understood in terms of normalization, which is a common neural mechanism found to operate in perceptual systems (Carandini & Heeger, 2012); for example, photoreceptor responses in the retina and neuron responses in primary visual cortex are normalized by dividing the responses by the response of the surrounding receptors or neurons (corresponding to the prevailing mean luminance or mean contrast, respectively). The normalization process optimizes the dynamic operation ranges of receptors and neurons to a variety of environmental conditions. The normalization model has been suggested to explain several visual-attention effects: For example, due to normalization, sensitivity to low-contrast target stimuli is increased and the effect of distractors is reduced (Reynolds & Heeger, 2009). Recently, it has been shown that noise in a neural model containing divisive normalization and population coding can account for errors in memory precision (Bays, 2014). For cross-modal representations, working memory could operate by normalizing resources across representations within the cognitive map: Each working memory representation is always allocated $1/n$ of the total resources, and thus is stored at the expense of other representations on the cognitive map.

In conclusion, our results suggest that working memory resources are shared across sensory modalities and that the precision of memory representations determines the amount of information that can be retained at a time. For every tested

visual and auditory feature, the resource-sharing function fit the data very well. Resource sharing could be understood as normalization of resources across memory representations, in order to optimize the usage of limited capacity.

Author Note This work was supported by the Academy of Finland (Grant No. 260054). We thank Daryl Fougnie and two anonymous reviewers for their helpful comments.

References

- Anderson, D. E., & Awh, E. (2012). The plateau in mnemonic resolution across large set sizes indicates discrete resource limits in visual working memory. *Attention, Perception, & Psychophysics*, *74*, 891–910. doi:10.3758/s13414-012-0292-1
- Anderson, D. E., Vogel, E. K., & Awh, E. (2011). Precision in visual working memory reaches a stable plateau when individual item limits are exceeded. *Journal of Neuroscience*, *31*, 1128–1138. doi:10.1523/JNEUROSCI.4125-10.2011
- Baddeley, A. (1986). *Working memory*. Oxford, UK: Oxford University Press, Clarendon Press.
- Baddeley, A. (2010). Working memory. *Current Biology*, *20*, R136–R140. doi:10.1016/j.cub.2009.12.014
- Baddeley, A. D., & Hitch, G. J. (1974). Working memory. In G. H. Bower (Ed.), *The psychology of learning and motivation: Advances in research and theory* (Vol. 8, pp. 47–89). New York, NY: Academic Press.
- Bays, P. M. (2014). Noise in neural populations accounts for errors in working memory. *Journal of Neuroscience*, *34*, 3632–3645. doi:10.1523/JNEUROSCI.3204-13.2014
- Bays, P. M., Catalao, R. F., & Husain, M. (2009). The precision of visual working memory is set by allocation of a shared resource. *Journal of Vision*, *9*(10), 7. doi:10.1167/9.10.7. 1–11.
- Bays, P. M., & Husain, M. (2008). Dynamic shifts of limited working memory resources in human vision. *Science*, *321*, 851–854. doi:10.1126/science.1158023
- Brooks, L. R. (1968). Spatial and verbal components of the act of recall. *Canadian Journal of Experimental Psychology*, *22*, 349–368. doi:10.1037/h0082775
- Carandini, M., & Heeger, D. J. (2012). Normalization as a canonical neural computation. *Nature Reviews Neuroscience*, *13*, 51–62. doi:10.1038/nrn3136
- Cocchini, G., Logie, R. H., Della Sala, S., MacPherson, S. E., & Baddeley, A. D. (2002). Concurrent performance of two memory tasks: Evidence for domain-specific working memory systems. *Memory & Cognition*, *30*, 1086–1095. doi:10.3758/BF03194326
- Cowan, N. (1997). *Attention and memory*. New York, NY: Oxford University Press.
- Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, *24*, 87–114. doi:10.1017/S0140525X01003922. disc. 114–185.
- Cowan, N. (2011). The focus of attention as observed in visual working memory tasks: Making sense of competing claims. *Neuropsychologia*, *49*, 1401–1406. doi:10.1016/j.neuropsychologia.2011.01.035
- Degerman, A., Rinne, T., Pekkola, J., Autti, T., Jaaskelainen, I. P., Sams, M., & Alho, K. (2007). Human brain activity associated with audiovisual perception and attention. *NeuroImage*, *34*, 1683–1691. doi:10.1016/j.neuroimage.2006.11.019
- Fougnie, D., Asplund, C. L., & Marois, R. (2010). What are the units of storage in visual working memory? *Journal of Vision*, *10*(12), 27. doi:10.1167/10.12.27
- Fougnie, D., Cormiea, S. M., & Alvarez, G. A. (2013). Object-based benefits without object-based representations. *Journal of Experimental Psychology: General*, *142*, 621–626. doi:10.1037/a0030300
- Fougnie, D., & Marois, R. (2006). Distinct capacity limits for attention and working memory: Evidence from attentive tracking and visual working memory paradigms. *Psychological Science*, *17*, 526–534. doi:10.1111/j.1467-9280.2006.01739.x
- Fougnie, D., & Marois, R. (2011). What limits working memory capacity? Evidence for modality-specific sources to the simultaneous storage of visual and auditory arrays. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*, 1329–1341. doi:10.1037/a0024834
- Fougnie, D., Suchow, J. W., & Alvarez, G. A. (2012). Variability in the quality of visual working memory. *Nature Communication*, *3*, 1229. doi:10.1038/ncomms2237
- Franconeri, S. L., Alvarez, G. A., & Cavanagh, P. (2013). Flexible cognitive resources: Competitive content maps for attention and memory. *Trends in Cognitive Sciences*, *17*, 134–141. doi:10.1016/j.tics.2013.01.010
- Huang, L. (2010a). Visual working memory is better characterized as a distributed resource rather than discrete slots. *Journal of Vision*, *10*(14), 8. doi:10.1167/10.14.8
- Huang, L. (2010b). What is the unit of visual attention? Object for selection, but Boolean map for access. *Journal of Experimental Psychology: General*, *139*, 162–179. doi:10.1037/a0018034
- Kroll, N. E., Parks, T., Parkinson, S. R., Bieber, S. L., & Johnson, A. L. (1970). Short-term memory while shadowing: Recall of visually and of aurally presented letters. *Journal of Experimental Psychology*, *85*, 220–224. doi:10.1037/h0029544
- Kumar, S., Joseph, S., Pearson, B., Teki, S., Fox, Z. V., Griffiths, T. D., & Husain, M. (2013). Resource allocation and prioritization in auditory working memory. *Cognitive Neuroscience*, *4*, 12–20. doi:10.1080/17588928.2012.716416
- Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature*, *390*, 279–281. doi:10.1038/36846
- Luck, S. J., & Vogel, E. K. (1998). Response from Luck and Vogel. *Trends in Cognitive Sciences*, *2*, 78–79.
- Machizawa, M. G., Goh, C. C., & Driver, J. (2012). Human visual short-term memory precision can be varied at will when the number of retained items is low. *Psychological Science*, *23*, 554–559. doi:10.1177/0956797611431988
- Morey, C. C., & Cowan, N. (2004). When visual and verbal memories compete: Evidence of cross-domain limits in working memory. *Psychonomic Bulletin & Review*, *11*, 296–301. doi:10.3758/BF03196573
- Morey, C. C., & Cowan, N. (2005). When do visual and verbal memories conflict? The importance of working-memory load and retrieval. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *31*, 703–713. doi:10.1037/0278-7393.31.4.703
- Morey, C. C., Cowan, N., Morey, R. D., & Rouder, J. N. (2011). Flexible attention allocation to visual and auditory working memory tasks: Manipulating reward induces a trade-off. *Attention, Perception, & Psychophysics*, *73*, 458–472. doi:10.3758/s13414-010-0031-4
- Murray, A. M., Nobre, A. C., Astle, D. E., & Stokes, M. G. (2012). Lacking control over the trade-off between quality and quantity in visual short-term memory. *PLoS ONE*, *7*, e41223. doi:10.1371/journal.pone.0041223
- Oberauer, K., & Eichenberger, S. (2013). Visual working memory declines when more features must be remembered for each object.

- Memory & Cognition*, 41, 1212–1227. doi:10.3758/s13421-013-0333-6
- Olson, I. R., & Jiang, Y. (2002). Is visual short-term memory object based? Rejection of the “strong-object” hypothesis. *Perception & Psychophysics*, 64, 1055–1067. doi:10.3758/BF03194756
- Palmer, J. (1990). Attentional limits on the perception and memory of visual information. *Journal of Experimental Psychology: Human Perception and Performance*, 16, 332–350. doi:10.1037/0096-1523.16.2.332
- Reynolds, J. H., & Heeger, D. J. (2009). The normalization model of attention. *Neuron*, 61, 168–185. doi:10.1016/j.neuron.2009.01.002
- Salmela, V. R., Lähde, M., & Saarinen, J. (2012). Visual working memory for amplitude-modulated shapes. *Journal of Vision*, 12(6), 2. doi:10.1167/12.6.2
- Salmela, V. R., Mäkelä, T., & Saarinen, J. (2010). Human working memory for shapes of radial frequency patterns. *Vision Research*, 50, 623–629. doi:10.1016/j.visres.2010.01.014
- Salmela, V. R., & Saarinen, J. (2013). Detection of small orientation changes and the precision of visual working memory. *Vision Research*, 76, 17–24. doi:10.1016/j.visres.2012.10.003
- Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, 136, 663–684. doi:10.1037/0096-3445.136.4.663
- Scarborough, D. L. (1972). Stimulus modality effects on forgetting in short-term memory. *Journal of Experimental Psychology*, 95, 285–289. doi:10.1037/h0033667
- Sims, C. R., Jacobs, R. A., & Knill, D. C. (2012). An ideal observer analysis of visual working memory. *Psychological Review*, 119, 807–830. doi:10.1037/a0029856
- van den Berg, R., Shin, H., Chou, W. C., George, R., & Ma, W. J. (2012). Variability in encoding precision accounts for visual short-term memory limitations. *Proceedings of the National Academy of Sciences*, 109, 8780–8785. doi:10.1073/pnas.1117465109
- Van der Burg, E., Awh, E., & Olivers, C. N. (2013). The capacity of audiovisual integration is limited to one item. *Psychological Science*, 24, 345–351. doi:10.1177/0956797612452865
- Vergauwe, E., Barrouillet, P., & Camos, V. (2010). Do mental processes share a domain-general resource? *Psychological Science*, 21, 384–390. doi:10.1177/0956797610361340
- Wheeler, M. E., & Treisman, A. M. (2002). Binding in short-term visual memory. *Journal of Experimental Psychology: General*, 131, 48–64. doi:10.1037/0096-3445.131.1.48
- Wilken, P., & Ma, W. J. (2004). A detection theory account of change detection. *Journal of Vision*, 4(12), 1120–1135. doi:10.1167/4.12.11
- Xu, Y. (2002). Encoding color and shape from different parts of an object in visual short-term memory. *Perception & Psychophysics*, 64, 1260–1280. doi:10.3758/BF03194770
- Zhang, W., & Luck, S. J. (2008). Discrete fixed-resolution representations in visual working memory. *Nature*, 453, 233–235. doi:10.1038/nature06860
- Zhang, W., & Luck, S. J. (2011). The number and quality of representations in working memory. *Psychological Science*, 22, 1434–1441. doi:10.1177/0956797611417006