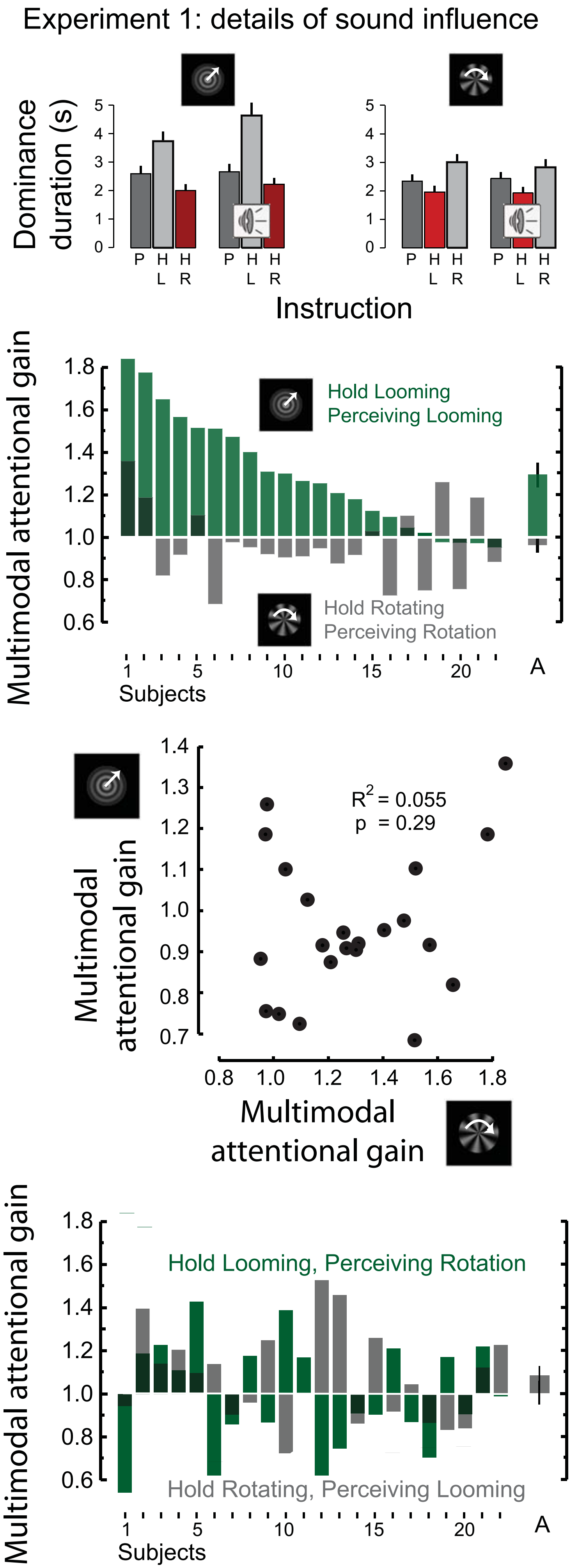


Fig Suppl1
van Ee *et al.*

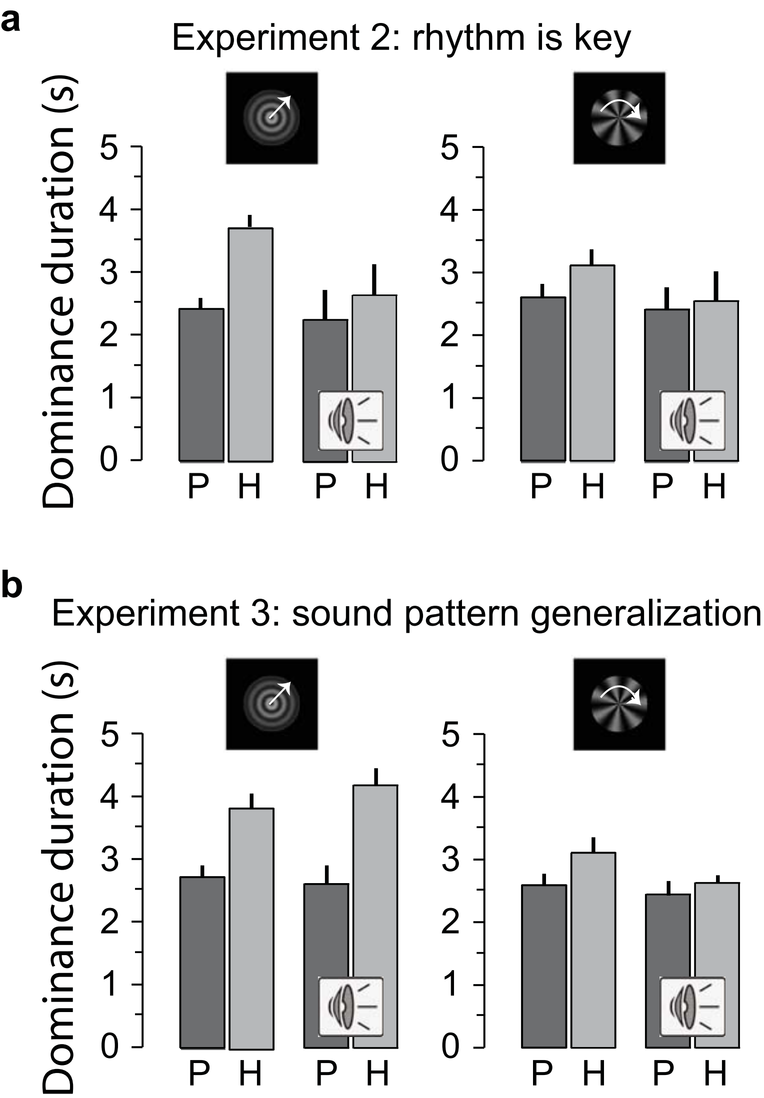


Fig Suppl2
van Ee *et al*.

Experiment 4: role of attention

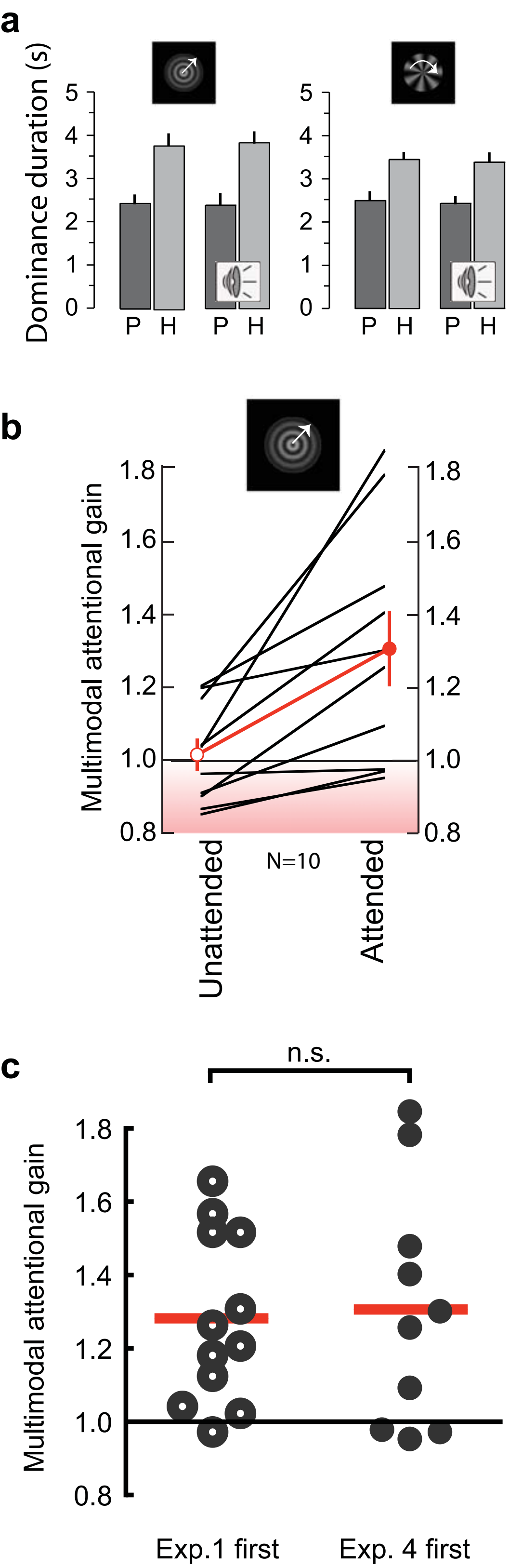


Fig Suppl3
van Ee *et al.*

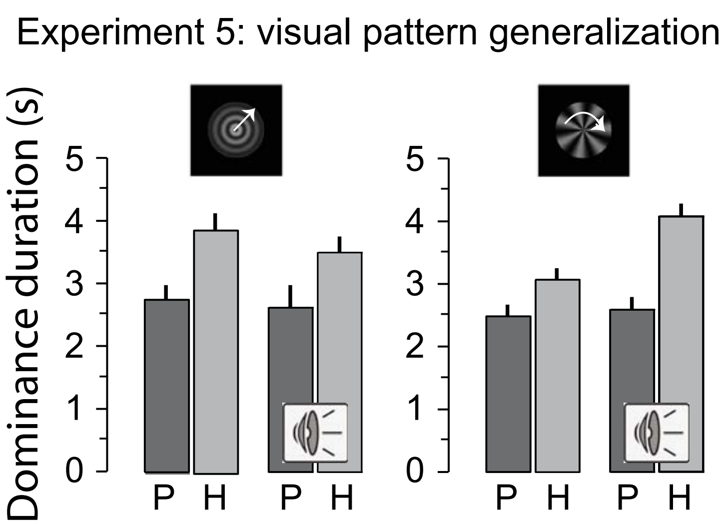
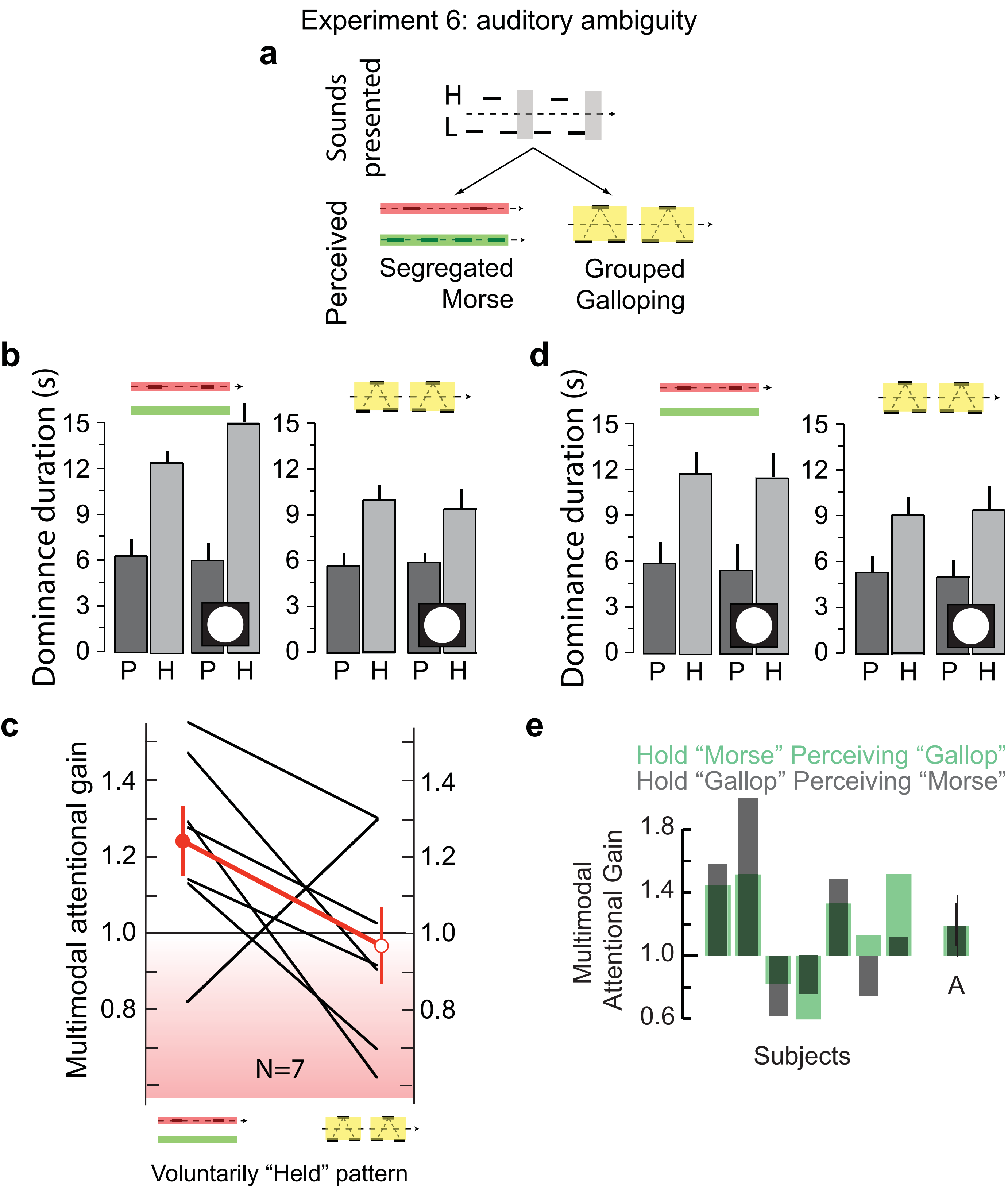


Fig Suppl4
van Ee *et al.*



Experiment 6: auditory ambiguity
a
Sounds presented
H
L
Perceived
Segregated Morse
Grouped Galloping
b
Dominance duration (s)
P H P H
d
c
Multimodal attentional gain
N=7
Voluntarily "Held" pattern
e
Hold "Morse" Perceiving "Gallop"
Hold "Gallop" Perceiving "Morse"
Multimodal Attentional Gain
A
Subjects

Experiment 7: touch and sound

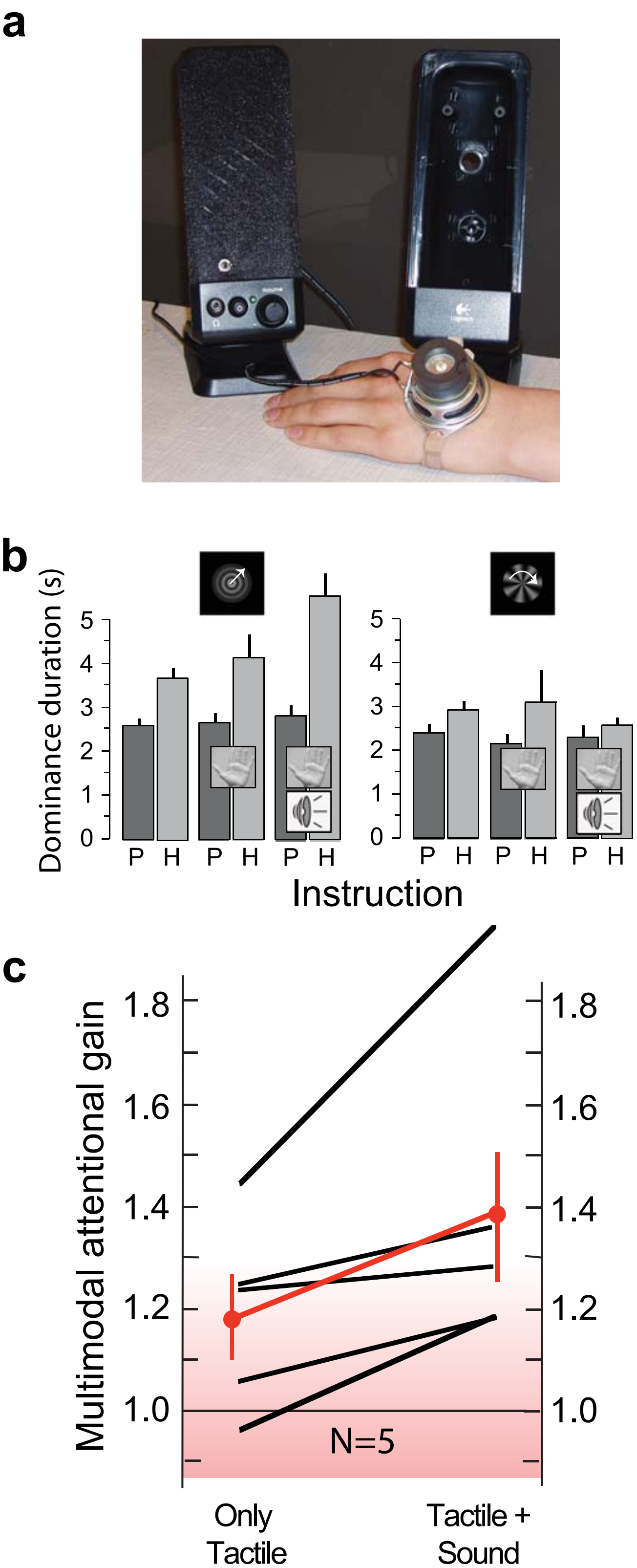


Fig Suppl6
van Ee *et al.*

Multisensory congruency as a mechanism for attentional control over perceptual selection

Raymond van Ee, Jeroen J.A. van Boxtel, Amanda L. Parker, David Alais

Supplement: Results in detail

Binocular rivalry tracking data for a total of 816 mins (n = 22) were collected to quantify the baseline influence of sound on attentional control of rivalry dominance. There were 6 conditions, with 16 subjects doing 4 min blocks and 6 subjects (who did three repetitions of the experiment) doing 12 min blocks. The mean perceptual durations for the passive and the two active 'hold' conditions are plotted in Figure Suppl1a (part of these data were presented in Fig 2), shown separately for the looming visual pattern (left-hand panel) and the radial visual pattern (right-hand panel). For the raw dominance durations of the looming pattern, the mean dominance duration without sound (first pair of columns) was 2.6±0.2 s in the passive condition, and subjects were able to increase dominance to 3.6±0.3 s by attentively holding it ($p<0.001$, paired t-test). For the radial pattern, mean dominance duration without sound (right panel, first pair of columns) was 2.3±0.2 s in the passive condition, increasing to 3.0±0.3 s when subjects attentively held it ($p<0.001$, paired t-test). These findings from the no-sound conditions replicate earlier reports showing a degree of voluntary attentional control in selecting the dominant percept in perceptual bistability. These findings also replicate earlier reports showing that not only the dominance periods for the held pattern change, but also those for the other pattern. When subjects held the looming pattern the dominance duration of the radial pattern changed slightly from 2.3±0.2 s to 1.9±0.2 s ($p<0.001$). When subjects held the radial pattern the dominance duration of the looming pattern changed from 2.6±0.2 s to 2.0±0.2 s ($p<0.001$). There were brief periods of superimposed or piecemeal pattern perception scattered throughout the observation period where neither pattern was exclusively dominant. In total, they averaged 13.7% of the observation period.

Next we turn to the multimodal conditions (see also Fig Suppl1a), in which the visual stimuli and attentional selection tasks were identical to those just described above, but there was a looming sound matched to the visual looming pattern. During passive viewing, mean dominance duration for the looming pattern was identical ($p>0.85$, paired t-test) to the passive-no-sound condition (2.5±0.3 s), but subjects were able to nearly double mean dominance duration to 4.5±0.4 s in the active sound condition by attentively holding it ($p<10^{-4}$). The absence of a change in dominance durations in the passive condition shows that presenting sound with the visual stimuli did not automatically change dominance durations. These findings were corroborated by a two-way repeated-measures ANOVA (sound (present/absent) x hold (yes/no)), revealing significant main effects for sound ($F_{1,21}$=10.4, $p<.005$) and hold ($F_{1,21}$=55.2, $p<0.005$). Importantly, there was a significant interaction between these factors ($F_{1,21}$=22.1, $p<0.001$), indicating that subjects were better able to hold the looming pattern with the sound present than with the sound absent.

For the radial pattern, similarly, the mean dominance duration in the passive condition was virtually unchanged by the presence of sound (2.3±0.2 s vs 2.3±0.2 s without sound; $p>0.80$, paired t-test), and subjects were able to increase it to 2.9±0.3 s by attentively holding it ($p<0.001$, paired t-test). A similar two-way ANOVA for the radial pattern revealed no effect of sound ($F_{1,21}$=0.35,$p>0.7$), but a significant increase of dominance durations when subjects attempted to influence dominance ($F_{1,21}$=20.4, $p<0.0001$). The interaction was not significant ($F_{1,21}$=1.2, $p=0.27$), although it showed a trend in the opposite direction to the looming pattern, suggesting that subjects were worse at holding the radial pattern when the looming sound was present. We also examined predominance of the looming and radial patterns and ruled out that sound had a non-specific enhancing effect: the non-held pattern duration did not increase with sound present (all $p>0.3$; paired t-test). In these multimodal conditions, the dominance duration of the radial pattern changed from 2.3±0.2 s (for passive) to 1.9±0.2 s (hold looming). When subjects held the radial pattern the dominance duration of the looming pattern changed from 2.5±0.3 s (passive) to 2.1±0.2 s (hold radial pattern). Again a two-way ANOVA for the non-held patterns showed only significant effects of the hold-condition ($p<0.0001$).

Detailed results of the other experiments are given in the figures.

Captions of Supplementary Figures

Fig Suppl1: Detailed data from Experiment 1 (part of the data were presented in Figs. 2a and 3a of the main paper). **a.** The mean duration for the looming (left) and the radial (right) visual patterns. P denotes passive; HL and HR denote hold the looming pattern and rotating pattern, respectively. The 'speaker icon' denotes the sound conditions. **b.** The proportion of the gain with the sound present over the gain without sound quantifies how much the looming sound enhances attentional control over the visual pattern. Those proportions for each individual subject (horizontally) are presented as multimodal attentional gain for the looming (green) and the radial (grey) patterns. The average value across the 22 subjects is denoted by A. **c.** The lack of correlation between the multimodal attentional gains for the two patterns demonstrates that a subject who is successful in holding the looming pattern is not necessarily successful in holding the radial pattern. **d.** Same as panel b, but now for the non-held patterns. Neither of them produces data significantly greater than 1 (p's>0.1). Error bars, ±1 standard error.

Fig Suppl2: The mean duration for the looming (left) and the radial (right) visual patterns in Experiment 2 (**a)** and Experiment 3 (**b)**. P denotes passive; H denotes hold. Error bars, ±1 standard error.

Fig Suppl3: **a.** The mean duration for the looming (left) and the radial (right) visual patterns in Experiment 4. Error bars, ±1 standard error. **b.** Direct comparison of the data from Experiment 1 (Fig. 3a of main text) and Experiment 4 (Fig. 3d) for each individual who ran both the attended and unattended conditions, demonstrating that paying attention to the congruent looming sound is required to enhance holding of the visual looming pattern. The dashed lines connect data of identical subjects. The filled red circle indicates significance (t-test, see text), and error bars are ±1 standard error. **c.** Multimodal attentional gain in Experiment 1, comparing subjects who started with Experiment 1 (open circles) with those who first completed Experiment 4 (closed circles). These data show that there is no significant difference between the two groups ($p>0.8$), while both groups show multimodal attentional gains larger than 1 ($p<0.002$, and $p<0.02$, respectively), implying that there is no effect of practice. These results show that the absence of multimodal attentional gain in Experiment 4 is unlikely to be due to inexperience of the subjects.

Fig Suppl4: The mean duration for the looming (left) and the radial (right) visual patterns in Experiment 5. Error bars, ±1 standard error.

Fig Suppl5: **a.** We presented a high-frequency pure tone (H) alternating with a low-frequency pure tone (L) in an LHL pattern. This sequence can be perceived either as one stream (LHL-LHL, i.e., grouped 'galloping' rhythm) or as two streams (H-H-H and L-L-L, i.e., segregated 'Morse' tones). **b.** The mean duration for the Morse (left) and the galloping (right) auditory patterns. Note that the scale is different as in the other mean durations data figures. The disk icon denotes the visual flickering disk that supported the Morse pattern. **c.**

There was significant multimodal gain in holding the percept of segregated Morse tones dominant over the galloping tones when the flickering visual disk (matching the Morse) was viewed, but there was no significant change when holding the galloping pattern. **d.** Same as panel b but here subjects were given the instruction that the flickering disk was not relevant to their task, although no explicit instruction was given to attend or to disregard the disk. These data concern a preliminary pilot experiment whose conditions were exactly identical to those used to collect the data for panel b but it involved only four out of the seven subjects who participated in this experiment (note that those four subjects participated in all experiments presented in this paper). The duration of a pilot serie was 2 minutes. They did this attention task before they participated in Expt 6. **e.** Same as panel d of Fig. Suppl1 but now for the 7 subjects whose data are plotted in panel c of this figure: changes in "multimodal attentional gain" for the non-held patterns. Neither of the two multimodal attentional gains is significantly greater than 1 (p's>0.2). The average value across the 7 subjects is denoted by A. Error bars show ±1 standard error.

Fig Suppl6: **a.** A sound speaker was attached to the dorsal side of the hand to produce a tactile looming stimulus matched to both the visual and the auditory looming stimuli. **b.** The mean duration for the looming (left) and the radial (right) visual patterns in tactile condition (middle pair of bars) and the tactile+sound condition (right pair of bars). The 'hand icon' denotes the tactile conditions. **c.** The individual subject data for both the tactile and the tactile+sound conditions. Error bars, ±1 standard error.