

# Bimodal sensory discrimination is finer than dual single modality discrimination

**Ansgar Koene**

Department of Psychology, University College London,  
London, UK,  
Knowledge Creating Communication Research Center,  
NICT, Kyoto, Japan, &  
ATR Computational Neuroscience Laboratories,  
Kyoto, Japan



**Derek Arnold**

School of Psychology, The University of Queensland,  
St. Lucia, Queensland, Australia



**Alan Johnston**

Departments of Psychology and Complex,  
University College London, London, U.K



Here we show that discriminating between different signal modulation rates can be easier when stimuli are presented in two modalities (vision and audition) rather than just one. This was true even when the single modality signal was repeated. This facilitation did not require simultaneous presentations in both modalities and therefore cannot rely on sensory fusion. Signal detection threshold for bimodal signals and double single modality signals were found to be equivalent indicating that the double single modality signals were not intrinsically noisier. The lack of facilitation in double single modality conditions was not due to inaccessibility of the first sample because there is no performance difference when noise was added to either the first or second samples. We propose that the bimodal signal discrimination advantage arises from fluctuations in the magnitude of sensory noise over time and because observers select the most reliable modality on a trial by trial basis. Noise levels within repeated single modality trials are more likely to be similar than those within signals from different modalities. As a consequence, signal selection would be less effective in the former circumstances. Overall, our findings illustrate the advantage of using separate sensory channels to achieve reliable information processing.

Keywords: bimodal, signal discrimination, optimal cue integration, audio, visual

Citation: Koene, A., Arnold, D., & Johnston, A. (2007). Bimodal sensory discrimination is finer than dual single modality discrimination. *Journal of Vision*, 7(11):14, 1–11, <http://journalofvision.org/7/11/14/>, doi:10.1167/7.11.14.

## Introduction

Many environmental events generate corresponding signals in multiple modalities. For instance, a passing honeybee provides an audiovisual signal that indicates both its location and direction of movement. Integration of information from multiple modalities can therefore potentially enhance the precision and accuracy of our sensory perception.

Cross modal interaction can also qualitatively alter stimulus appearance, as in the double flash illusion where two auditory tones heard around the same time as a single flash can create the subjective impression of a double visual flash (Shams, Kamitani, & Shimojo, 2000, 2002; Violentyev, Shimojo, & Shams, 2005). The temporal window for these audiovisual interactions has been estimated as being approximately 100 ms (Shams et al., 2002). However, the exact size of the temporal window for multisensory integration is likely to be stimulus specific because it has been shown that subjects are unable to detect stimulus

onset asynchrony for audiovisual stimulus pairs that are presented within 21 ms in the case of sound before light and within 150 ms for light before sound (Stone et al., 2001).

Audio beeps can also influence the number of reported tactile taps (Bresciani et al., 2005). This cross modal modulating affect of audition on tactile perception is reduced when the stimuli in the two modalities are presented asynchronously, implying that audio and tactile signals are only combined if they are likely to be generated by the same stimulus. Similarly, Bresciani, Dammeier, and Ernst (2006) have found that altering the numbers of flashes can bias the number of taps reported and vice versa. The strength of this effect depends on the reliability of the signals in the two modalities. However, in these counting experiments, it can be difficult to distinguish a perceptual bias from a response bias.

In the ventriloquist effect, sound is typically mislocated in the direction of a visual stimulus (Alais & Burr, 2004; Battaglia, Jacobs, & Aslin, 2003; Bertelson, Vroomen, de Gelder, & Driver, 2000; Vroomen, Bertelson, & de

Gelder, 2001). This effect relies on the fact that vision usually provides a more reliable cue to position than audition. Alais and Burr (2004) have shown that the apparent location of a discrepant audiovisual target depends on the sum of the audio and the visual locations weighted by the reliability of each signal. This reliability-based weighted summations rule, previously described for other mixed modality tasks (Ernst & Banks, 2002), can also increase sensitivity. For instance, Hillis, Ernst, Banks, and Landy (2002) found that bimodal stimulus discrimination was better than single modality discrimination for visual and haptic information, as would be predicted from a maximum likelihood strategy in which independent estimates are weighted in inverse proportion to the variance of the estimates (Ernst, 2006; Hillis et al., 2002; Landy, Maloney, Johnston, & Young, 1995; Yuille & Buelthoff, 1996). Bresciani et al. (2006) have also shown that the judgments of the number of stimuli presented can be more reliable when two modalities provided the same information compared to just one. However, this advantage was only significant when subjects reported on the tactile cue—the more reliable cue.

Improvements in sensory discrimination can provide reliable evidence of optimal cue combination, even in situations when the signals carry equal value and are equally discriminable. However, if we simply compare bimodal stimuli with single modality cues, the bimodal benefits could be attributed to differences in the amount of information provided. We therefore decided to compare bimodal presentation against both single modality cues and double single modality cues. This necessitates the additional control of sequential as well as simultaneous bimodal presentations. This manipulation also allowed us to test whether any benefits in discrimination depend on stimulus simultaneity.

We chose to investigate how audio and visual perceptual cues are combined to detect and to judge differences in an amodal stimulus property, temporal frequency. We first compared bimodal and repeated unimodal conditions in a signal detection task to determine whether there was any evidence of effective summation of the bimodal signals at threshold, which would be indicative of sensory fusion. Judgments were based on (1) single visual or (2) auditory stimuli, (3) simultaneous or (4) sequential audiovisual stimuli, and (5) repeated visual or (6) auditory stimuli. The (1) single visual and (2) single auditory stimulus conditions provided baseline performance data.

## Experiment 1: Signal detection

The signal detection experiment measured the signal to noise ratios at which subjects are able to detect the presence of a periodic (10 Hz) amplitude/intensity modulation embedded in band-limited white noise. If bimodal facilitation of signal detection requires perceptual fusion, this facilitation should occur only in the simultaneous

bimodal condition (Condition 3). If, on the other hand, bimodal facilitation of signal detection is simply a consequence of an increase in stimulus information, then we predict that both bimodal conditions (Conditions 3 and 4) as well as the repeated unimodal conditions (Conditions 5 and 6) should show the same facilitation. We therefore compared signal detection performance in the bimodal conditions (Conditions 3 and 4) and in the repeated unimodal conditions (Conditions 5 and 6) against the baseline performances (Conditions 1 and 2) and performance predicted on the basis of “probability summation” (Macmillan & Creelman, 2005).

## Method

### Participants

Participants were one of the authors and three paid volunteers who were naive as to the purpose of the experiment. All had normal or corrected-to-normal vision and normal hearing.

### Apparatus and stimuli

Visual stimuli were presented on a 19" SONY Trinitron Multiscan 500PS monitor (160 Hz frame rate) and were generated using a Cambridge Systems VSG card controlled with Matlab 6 (the MathWorks) running on a Pentium 4 PC with Windows 2000 operating system. Auditory stimuli were generated using a Tucker Davis Technologies (TDT) RP2.1 Enhanced Real-Time Processor. For the simultaneous bimodal condition, onset times of the auditory and the visual stimuli were synchronized using trigger outputs from the VSG card. Responses were given by means of a Cambridge Systems CB3 experiment response box. The participant sat in a quiet room 57 cm from the stimulus monitor with background illumination. The visual stimulus consisted of a single luminance-modulated Gaussian blob ( $SD: 5^\circ$ ) in the center of a black screen with a central bull's-eye fixation point (diameter:  $0.2^\circ$ ) superimposed (Figure 1A). The luminance increment of the blob was temporally modulated between 0 and  $\sim 5$  cd/m<sup>2</sup>. A Gaussian blob modulated at 10 Hz provides a strong visual signal (Kelly, 1979). Participants were instructed to maintain fixation at the fixation point.

The auditory stimulus was a 3-kHz tone diotically presented via headphones (Sennheiser HDA 200) with a sampling frequency of 24420 Hz. The amplitude of the auditory tone was modulated in time (56 dB sound pressure level [SPL] at the peak of modulation).

All stimulus waveforms were amplitude modulated by a temporal Gaussian envelope ( $\sigma = 0.28$  s) to avoid sudden signal onset effects. The amplitude/luminance of the audio and the visual test stimuli was sinusoidally modulated at a rate of 10 Hz (Figure 1B). The S/N of the test stimulus was adjusted by adding band-limited ( $F_c = 10$  Hz,  $B_w =$

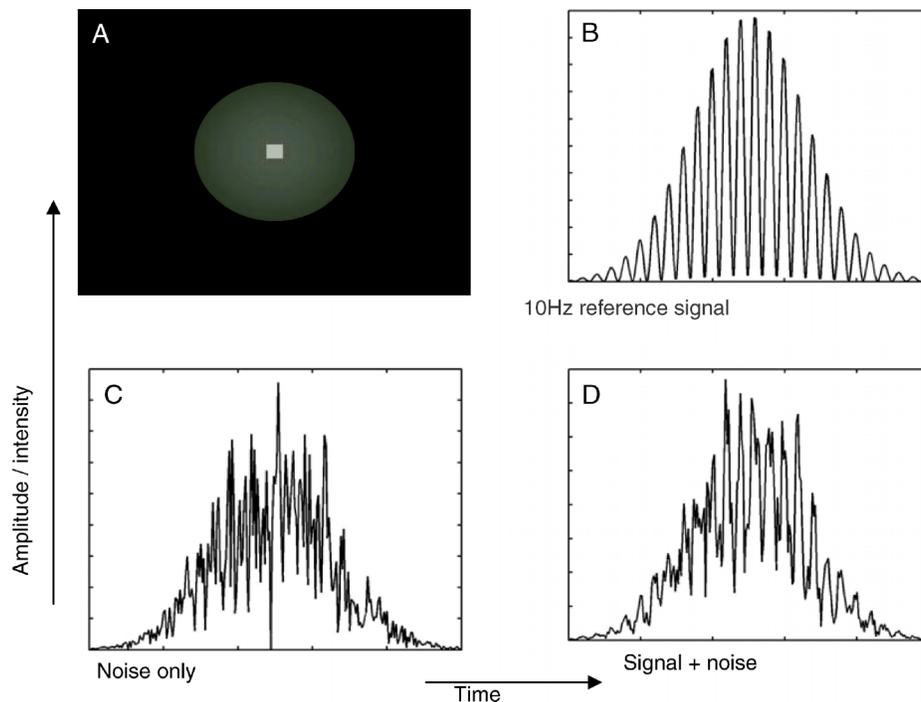


Figure 1. Panel A shows a schematic snapshot of the visual stimulus. Panels B, C, and D illustrate the temporal amplitude/intensity modulation of the audio and the visual stimuli (time on the X-axis and sound amplitude/light intensity on the Y-axis). Panel B shows the waveform of the 10-Hz reference signal. Panel C illustrates the waveform of the band-limited white noise used as the ‘noise-only’ comparison stimulus. Panel D illustrates the waveform of the test stimuli (i.e.,  $D = B' + C'$ , where  $B'$  is of type B with a different oscillation frequency and  $C'$  is of type C with independent random fluctuations).

5 Hz) white noise (Figure 1D). The same type of noise was used for the “noise-only” comparison stimulus (Figure 1C). The noise for both signals was independently sampled. A single stimulus presentation lasted 2.5 s.

### Procedure

We used a two-interval forced-choice design. For the “unimodal single presentation” and the “simultaneous bimodal” conditions, subjects were presented with the test signal (noise +10 Hz modulation) once per trial. This was either preceded or followed by a single presentation of the “noise-only” stimulus. For the “unimodal repeated presentation,” the test and the “noise only” stimuli were presented twice (immediate repetition with independent noise in both presentations) before the other stimuli were presented. The “sequential bimodal” condition followed the same procedure as the “repeated unimodal” conditions except that the stimulus repetition was in a different modality. The order of audio and visual modality presentation was randomized. At the end of each trial subjects indicated, by means of a button press, which interval contained the 10-Hz amplitude modulation.

Using a blocked design, we first determined detection thresholds for the audio and the visual “unimodal single presentation” conditions (randomly mixed blocks) using the method of constant stimuli with six different S/N

values (vision: 0.2, 0.4, 0.6, 0.8, 1, and 1.2; sound: 0.04, 0.06, 0.08, 0.1, 0.12, and 0.14, where S/N is defined as amplitude of sinusoidal signal component divided by peak amplitude of noise signal component). Each S/N level was repeated on 10 trials per session in a random mixed order with six sessions per condition. Detection thresholds were defined as the S/N for which observers gave 75% correct responses (determined using maximum likelihood Weibull function fits).

Subsequently, the simultaneous and the sequential bimodal conditions, as well as the “repeated unimodal” audio and the visual conditions, were tested in randomly mixed blocks. The S/N values for the test signals in the bimodal and the “repeated unimodal” conditions were chosen from 12.5%, 25%, 50%, 75%, 100%, 125%, and 150% of the detection thresholds for the “unimodal single presentation” conditions (Figure 2A). Because the S/N for the audio and the visual stimuli in the bimodal conditions are independent variables, there is a two-dimensional parameter space. Our goal was only to establish if signal detection performance for bimodal stimuli requires sensory fusion of the audio and the visual stimuli into a percept of a single event and, if not, whether repeated unimodal stimulation will achieve the same degree of facilitation as bimodal stimulation does. We therefore confined our experiment to stimuli with S/N for the auditory and the visual modalities that were equal

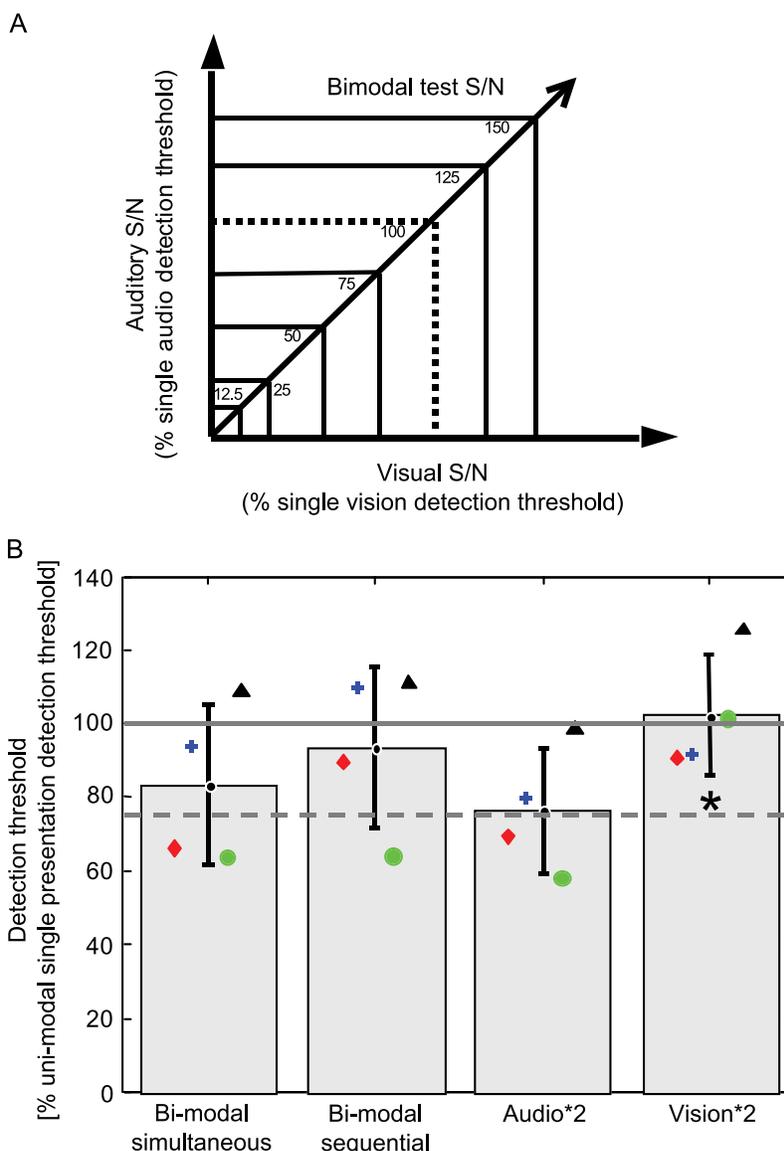


Figure 2. Panel A illustrates the choice of audio and visual S/N levels that were used in the bimodal and the repeated unimodal signal detection tasks. Along the axes are the S/N for the visual (X-axis) and auditory (Y-axis) stimuli. The dotted lines indicate the auditory and the visual detection thresholds. Test stimuli were chosen at 12.5–150% of the single presentation unimodal signal detection thresholds (thin solid lines). Panel B shows the detection thresholds (mean and standard deviation from four subjects) for the bimodal and the “repeated unimodal” conditions as a percentage of the detection thresholds for the “single presentation unimodal” conditions (solid line). The dashed line indicates the predicted detection threshold for the “repeated unimodal” and bimodal conditions based on probability summation (Pirenne, 1943; Treisman, 1998). The symbols indicate the detection thresholds for each of the four subjects (red  $\diamond$  = RJ, green  $\bullet$  = AK, black  $\blacktriangle$  = AA, and blue + = AB).

percentages of their respective unimodal detection thresholds (Figure 2A).

**Results and discussion**

Figure 2B shows the signal detection thresholds (mean and standard deviation) for the bimodal and the “repeated unimodal” conditions as a percentage of the detection thresholds for the “single presentation unimodal” conditions. The dashed line indicates the predicted

detection threshold for the “repeated unimodal” and bimodal conditions based on probability summation (Macmillan & Creelman, 2005; Pirenne, 1943; Treisman, 1998) with equal baseline detection thresholds for vision and audition.

A one-way ANOVA found no significant difference between the bimodal and the repeated unimodal conditions ( $df = 3, F = 1.382, p = .295$ ). One-sample  $t$  tests comparing the detection thresholds for the “repeated unimodal” and the bimodal conditions against the “single

presentation unimodal” detection thresholds (solid line at 100%) found no significant differences (bimodal simultaneous:  $df = 3$ ,  $t = 1.548$ ,  $p = .219$ ; bimodal sequential:  $df = 3$ ,  $t = 0.594$ ,  $p = .594$ ; repeated unimodal audio:  $df = 3$ ,  $t = 2.745$ ,  $p = .071$ ; repeated unimodal vision:  $df = 3$ ,  $t = 0.286$ ,  $p = .794$ ). Comparison against the prediction from probability summation only showed a significant difference for the “repeated unimodal vision” condition (bimodal simultaneous:  $df = 3$ ,  $t = 0.756$ ,  $p = .505$ ; bimodal sequential:  $df = 3$ ,  $t = 1.7$ ,  $p = .188$ ; repeated unimodal audio:  $df = 3$ ,  $t = 0.164$ ,  $p = .88$ ; repeated unimodal vision:  $df = 3$ ,  $t = 3.371$ ,  $p = .043$ ).

Signal detection performance was somewhere between that predicted by probability summation and the performance of a single presentation in either of the two modalities (Treisman, 1998).

The results of our signal detection experiment showed no clear difference between either of the bimodal conditions (Conditions 3 and 4) or between the bimodal and the repeated unimodal conditions (Conditions 5 and 6). There was therefore no evidence of sensory signal summation because there were no benefits in the bimodal simultaneous case over the other conditions that could not be accounted for by probability summation.

From the results of Experiment 1, we can conclude that, in this context, the provision of a multisensory audiovisual signal did not aid 10 Hz signal detection relative to repeated audio or repeated visual signal presentations. However, it has been previously shown that the combination of audio and visual signals can facilitate discrimination for superthreshold signals (Alais & Burr, 2004). We therefore decided to examine discriminations between differing superthreshold temporal frequency signals.

## Experiment 2: Signal discrimination with equalized unimodal performance

In this signal discrimination experiment, we measured temporal frequency discrimination thresholds for amplitude/luminance intensity modulations of a periodic audio/visual signal, with respect to a reference modulation rate of 10 Hz. If bimodal facilitation of signal discrimination requires perceptual fusion, we predict that this facilitation should occur only in the simultaneous bimodal condition (Condition 3). If, on the other hand, bimodal facilitation is simply a consequence of an increase in stimulus information, we predict that both bimodal conditions (Conditions 3 and 4), as well as the repeated unimodal conditions (Conditions 5 and 6), should show the same facilitation. We therefore compared signal discrimination performance in the bimodal conditions (Conditions 3 and 4), the repeated unimodal conditions (Conditions 5 and 6), and

the baseline performances (Conditions 1 and 2) in addition to performance predicted by maximum likelihood optimal cue combination (Ernst & Banks 2002; Landy et al., 1995; Yuille & Buelthoff, 1996).

## Method

### Participants

Participants were two of the authors and two volunteers who were practiced observers but were naive as to the purpose of the experiment. All had normal or corrected-to-normal vision and normal hearing. Informed consent was obtained after the nature and the possible consequences of the study were explained.

### Apparatus and stimuli

The apparatus and the stimuli were the same as in Experiment 1 with the following modifications. The 10-Hz modulated audio and/or visual reference stimulus was presented without noise (Figure 1B). The test stimuli consisted of sinusoidal modulations with modulation rates of 7.5, 8.5, 9.5, 10.5, 11.5, and 12.5 Hz to which band-limited ( $F_c = 10$  Hz,  $B_w = 5$  Hz) white noise was added (Figures 1C and 1D). In the bimodal conditions, the sinusoidal modulation rate for the audio and the visual stimuli was identical. Subjects were informed that this would always be the case.

### Procedure

The procedure for the signal discrimination experiment employed a variant of the method of single stimuli and a blocked design in which each session contained trials from a single stimulus condition. The 10-Hz reference was presented 10 times at the beginning of each session in the modality tested in that session. Throughout the rest of the session only the test stimuli were presented. After presentation of a test stimulus, subjects were instructed to respond (by button press) if the sinusoidal modulation frequency of the stimulus was higher or lower than the 10-Hz reference modulation frequency. Symbolic feedback (vertical bar or high tone to indicate modulation rates higher than 10 Hz, horizontal bar or low tone to indicate modulation rates lower than 10 Hz) was given after each response. Test stimuli were selected in a pseudorandom order. Each modulation rate was presented 10 times per session. Each subject completed four sessions per condition. The six conditions were randomly mixed.

To avoid learning effects during the main experiment and to equate the S/N of the audio and the visual stimuli, subjects were trained in the “unimodal single presentation” conditions over a number of days until their signal discrimination performance stabilized. S/N levels for the audio and the visual stimuli were experimentally adjusted for each subject such that performances for the audio and

the visual “unimodal single presentation” conditions were approximately equal and in a range where any possible bimodal facilitation effects would be detectable. Once these S/Ns were established, they remained constant throughout the experiment.

Psychometric functions were fitted using the maximum-likelihood Probit-function fitting procedure adapted from

David Foster’s bootstrap program (Foster & Bischof, 1987) with 40 measurements per data point.

**Results and discussion**

Figure 3 shows discrimination thresholds for four subjects for each of the six experiment conditions

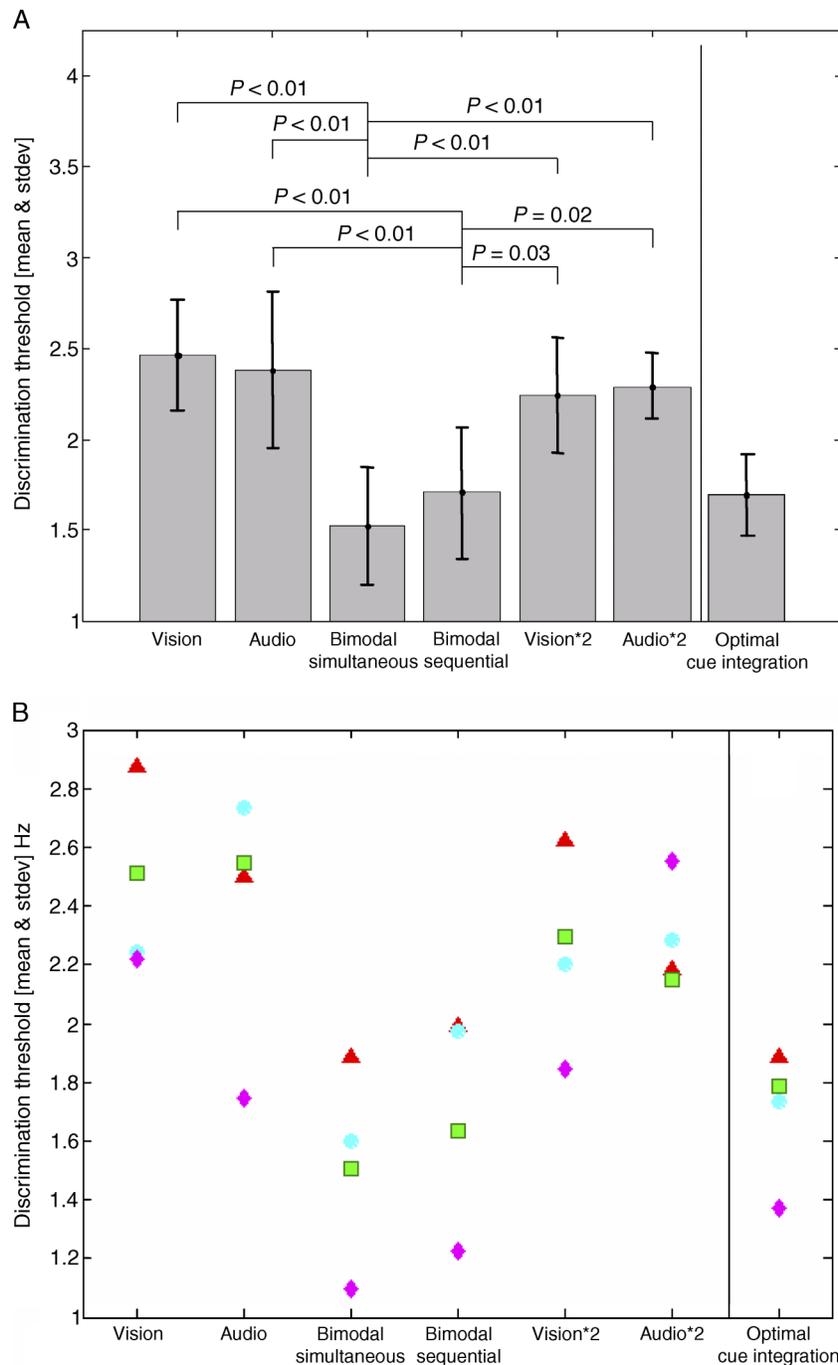


Figure 3. (A) Mean and standard deviation of discrimination thresholds of four subjects for the six conditions used in Experiment 2 and the “optimal cue integration” prediction for the bimodal conditions. The p values in panel A indicate all significant differences between stimulus conditions as determined by LSD post hoc tests. The error bars show standard deviations between subjects. Optimal cue integration predictions were calculated per subject, based on minimum variance optimization, and were subsequently averaged across subjects. (B) Individual subject data (red ▲ = JW, light blue ● = AB, magenta ◆ = DA, and green ■ = AK).

(Figure 3A: mean data; Figure 3B: individual data). A one-way ANOVA showed a significant effect of experimental condition across subjects ( $N = 4$ ,  $df = 5$ ,  $F = 5.553$ ,  $p = .003$ ). LSD post hoc tests (see Figure 3A) revealed a significant difference between bimodal conditions versus unimodal conditions, but no significant differences between sequential and simultaneous bimodal conditions or between unimodal single and repeated presentation conditions. There was no significant difference between the unimodal audio and the visual conditions.

The reduced discrimination threshold in the simultaneous bimodal condition indicates that observers are able to take advantage of the signals in both modalities when presented at the same time, confirming previous findings in conditions where perceptual sensory fusion can occur. However, the reduced discrimination threshold in the sequential bimodal condition indicates that the facilitation for bimodal stimuli does not require sensory fusion. Sensory information from multiple modalities can be tapped to facilitate discrimination of amodal signal properties, such as temporal modulation frequency, even over an extended period of time. This bimodal facilitation of signal discrimination is consistent with predictions for optimal cue combination (Ernst & Banks, 2002). Note that the optimal prediction for within-modal and intermodal tasks is the same here because single presentation audio and single presentation visual performances were equated. The optimal prediction is therefore the single presentation discrimination threshold divided by  $\sqrt{2}$ .<sup>1</sup>

The lack of significant facilitation for the “unimodal repeated presentation” conditions, relative to the “single presentation conditions,” raises the question of why observers were unable to take advantage of the extra information provided by the unimodal stimulus repetitions. It is possible that, because both stimuli were in the same modality, the second stimulus presentation may have interfered with recollection of the first. Alternatively, some kind of adaptation induced by the first stimulus presentation may have interfered with the full utilization of the second. To test these possibilities, we performed a second signal discrimination experiment aimed at determining if both the first and the second stimulus presentations were fully accessible to subjects during the bimodal sequential and the repeated unimodal trials.

## Experiment 3: Signal discrimination with different S/N for first and second stimulus

We tested if subjects were able to utilize the information in both the first and the second stimulus presentations

by using different S/Ns for the first and the second stimuli within a trial. If subjects are only able to use the information from the first stimulus presentation, performance should be best when the S/N is high in the first stimulus presentation and low in the second. Conversely, if subjects are only able to use the information from the second presentation, performance should be best when the S/N is low in the first stimulus presentation and high in the second. If, however, performances in the “low S/N first” and “high S/N first” cases are not significantly different, then there would be no evidence for an information bottleneck limiting the use of information in either stimulus presentation.

## Method

### Participants

There were two participants, one of the authors and a volunteer who was naive as to the purpose of the experiment. Both subjects had previously participated in Experiments 1 and 2.

### Apparatus and stimuli

The apparatus and the stimuli were the same as in Experiment 2 with the following modifications: “high S/N” stimuli had the same noise amplitudes that had been used for these subjects in Experiment 2; “low S/N” stimuli had noise amplitudes  $\sim 1$  order of magnitude greater. This was simply achieved by introducing a proportional change in the amplitude of the added noise.

### Procedure

The experimental procedure was the same as in Experiment 2, with the exception that the first and the second stimulus in the repeated unimodal and the sequential bimodal conditions had different S/N levels. The simultaneous bimodal condition was not tested because we were interested in temporal sequence effects. The four unimodal single presentation conditions (audio “high S/N,” audio “low S/N,” vision “high S/N,” and vision “low S/N”) were tested in random order prior to testing the bimodal and the “repeated unimodal” conditions.

Once the baseline measures were completed, the two “sequential bimodal” conditions (“first high S/N” followed by “low S/N” presentation and “first low S/N” followed by “high S/N” presentation) and the four “repeated unimodal” conditions (“first high S/N” followed by “low S/N” presentation and “first low S/N” followed by “high S/N” presentation for audio and vision) were tested in randomized blocks. Each session consisted of “first high S/N” and “first low S/N” trials randomly mixed with 10 trials per modulation rate for each ( $2 \times 5 \times 10 = 100$  trials per block session). Each condition was repeated

five times. In the sequential bimodal condition, half the “first high S/N” trials and half the “first low S/N” trials” were “first audio, then vision” trials, whereas the other half were “first vision, then audio” trials. To simplify the comparison, the “first audio then vision” trials and the “first vision then audio” trials were pooled together and were analyzed only according to S/N sequence.

### Results and discussion

Figure 4 shows discrimination thresholds (mean and standard deviation) for five sessions completed for each condition by both subjects. These data were obtained as in Experiment 2, with  $5 \times 10$  measurements per data point. Comparison of the “first high S/N” against the “first low S/N” conditions within each stimulus modality condition (audio, visual, and bimodal) revealed no significant differences (two-tailed independent samples  $t$  tests,  $df = 8$ :  $2^*V$  [AB]:  $t = 1.563$ ,  $p = .157$ ;  $2^*V$  [AK]:  $t = -0.242$ ,  $p = .815$ ;  $2^*S$  [AB]:  $t = 1.397$ ,  $p = .189$ ;  $2^*S$  [AK]:  $t = 1.004$ ,  $p = .345$ ; Seq Bi [AB]:  $t = 0.652$ ,  $p = .533$ ; Seq Bi [AK]:  $t = 2.056$ ,  $p = .074$ ). However, the difference between the “high S/N” and the “low S/N” trials was

highly significant for both the audio and the visual “single presentation unimodal” conditions (two-tailed independent samples  $t$  tests,  $df = 8$ : V [AB]:  $t = 3.098$ ,  $p = .015$ ; V [AK]:  $t = 4.030$ ,  $p = .004$ ; S [AB]:  $t = 3.067$ ,  $p = .015$ ; S [AK]:  $t = 5.346$ ,  $p = .001$ ).

The sequential bimodal and the “repeated unimodal” conditions were not significantly different from the “high S/N” “unimodal single presentation” conditions. This indicates that subjects were able to utilize the information from the “high S/N” stimulus regardless of whether it was presented first or second in the “repeated unimodal” condition. The significant difference between “sequential bimodal” and “repeated unimodal” conditions found in Experiment 2 was not found here. Note that elevating the noise in one of the two samples eliminates the bimodal benefit. These results are consistent with maximum likelihood-based optimal cue integration (Andersen, Tiippana, & Sams, 2005; Hillis et al., 2002; Hillis, Watt, Landy, & Banks, 2004; Landy et al., 1995; Yuille & Buelthoff, 1996), which would predict that greatly reducing the S/N in one of the two stimulus presentations would cause performance to be based almost entirely on the high S/N stimulus.

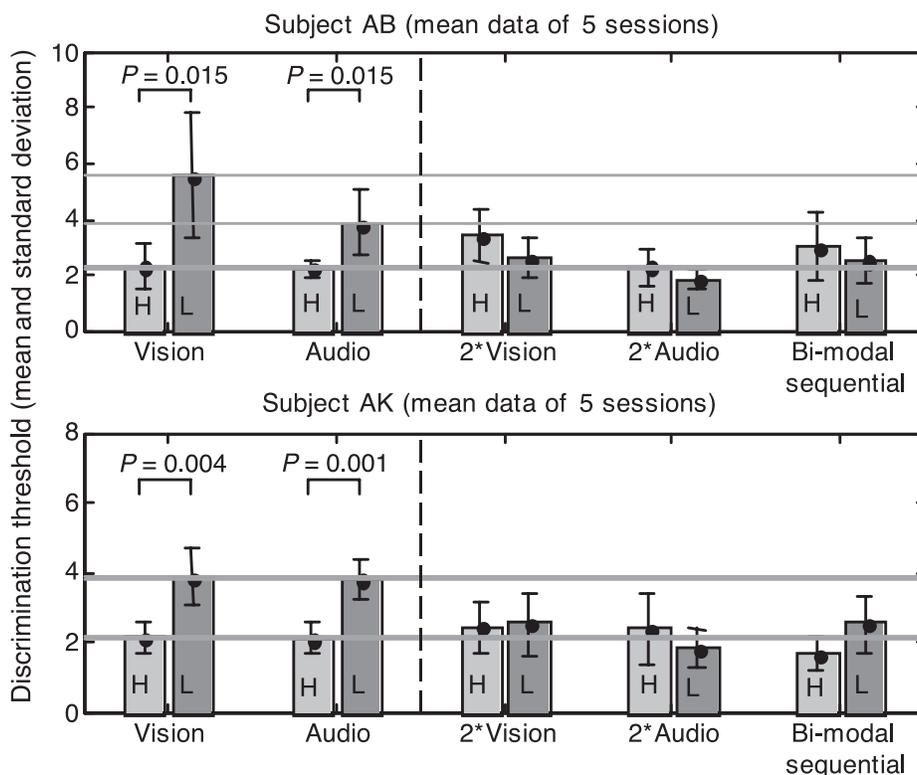


Figure 4. Top and bottom panels show the discrimination thresholds (mean and standard deviation) for the two subjects for the 10 conditions in Experiment 3 ( $5 \times 10$  trials per data point). The bars on the left of the dashed line show the “baseline” performance for the “high S/N” (light grey “H”) and the “low S/N” (dark grey “L”) “unimodal single presentation” conditions. The bars on the right of the dashed line show the results from the sequential bimodal and the “repeated unimodal” (audio and visual) conditions (light grey “H” = “first high S/N,” dark grey “L” = “first low S/N”). The  $p$  values indicate the results of two-tailed independent samples  $t$  tests.

## General discussion

We compared bimodal temporal frequency detection and discrimination against dual single modality signals. For signal detection ([Experiment 1](#)), we found no significant difference between the bimodal and the unimodal conditions or between the simultaneous and the sequential bimodal conditions. In short, we found no evidence for sensory signal integration over what could be accounted for by probability summation.

For signal discrimination ([Experiment 2](#)), we found significant facilitation not only for the simultaneous bimodal condition but also for the sequential bimodal condition, indicating that the bimodal advantage did not require perceptual fusion of the sensory inputs. Surprisingly, repeated stimulation in the same modality did not significantly improve discrimination performance over single presentation, although a control experiment ([Experiment 3](#)) showed that observers were able to utilize the information from both the first and the second unimodal stimulus presentations. The results of [Experiment 3](#) further revealed that the bimodal facilitation does not occur if the perceived S/N in the two modalities is greatly different.

### Attention effects

When comparing subject performance for different stimulus conditions, each presented in separate blocks, there is an inherent possibility that subject performance might be affected by differences in perceptual load. The simultaneous bimodal condition, during which subjects must attend to both audio and visual stimuli at the same time, would seem to be the most attention demanding condition. Performance for the sequential and the simultaneous bimodal conditions, however, was not significantly different—indicating that there was no disadvantage in having to process both the audio and the visual stimuli at the same time.

### Bimodal advantage or unimodal disadvantage

Experiments that have measured reaction times (RTs) to unimodal and bimodal stimuli have typically found a bimodal advantage, delivering shorter RTs for bimodal stimuli than for unimodal stimuli (e.g., Miller, 1982). As Patching and Quinlan (2004) have pointed out, however, the difference in RTs might result from a unimodal disadvantage, for example, modality-switching costs that may be present on unimodal trials (Quinlan & Hill, 1999; Spence & Driver, 1997) that require the subject to switch attention to a specific stimulus modality to respond. The blocked presentation of the stimulus conditions in our experiment makes it unlikely that modality-switching costs play a role. [Experiment 3](#) showed that the lack of

discrimination improvement in the “repeated unimodal” conditions was not due to an inability to use information from both stimulus presentations.

### Sensory fusion and bimodal discrimination facilitation

It has been previously shown that, under conditions supporting sensory fusion (i.e., near simultaneous multimodal stimulation), strong multimodal interaction effects can take place (Alais & Burr, 2004; Ernst & Banks, 2002; Patching & Quinlan, 2004; Shams et al., 2000, 2002; Violentsev et al., 2005). These include facilitation of both stimulus detection and discrimination (Bresciani et al., 2006; Hillis et al., 2002). Such multisensory interactions can occur even when subjects are expressly instructed to attend only to one modality and to ignore stimuli in the other modality. Here we showed, however, that sensory fusion conditions are not required for bimodal facilitation of temporal frequency discrimination. Performance in both the simultaneous and the sequential bimodal conditions was statistically indistinguishable from discrimination performance predicted by maximum likelihood estimation (Alais & Burr, 2004; Ernst & Banks, 2002; Hillis et al., 2004; Landy et al., 1995; Yuille & Buelthoff, 1996). This suggests that simultaneous bimodal activations of multimodal neurons are not essential for this kind of temporal frequency discrimination facilitation.

### Why is there no significant discrimination benefit in repeated unimodal conditions?

Considering that bimodal signal discrimination benefit does not require sensory fusion and the information content in the bimodal and dual single modality presentations is the same, why is there no significant benefit for dual single modality discrimination?

In [Experiment 3](#), in which the repeated unimodal signals had clearly different signal to noise ratios, subjects had no difficulty in basing their response on the signal with the greater S/N (see [Figure 4](#)); thus, it is not due to an inability to use both stimulus presentations. Indeed, the possibility of choosing the most reliable signal may provide the explanation.

Signal discrimination is limited by internal noise. Sources of internal noise are various, but we note that neuromodulators such as acetylcholine, which has been linked to the “reporting of uncertainty” (Dayan & Yu, 2002), are known to have onset time constants between 1 and 2 s and decay constants of 10 to 20 s (Hasselmo & Fehrlau, 2001). If the magnitude of the internal sensory noise in each modality changes slowly, and relatively independently, but significantly varies over the course of an experimental session, then there is an opportunity to improve performance in

bimodal trials by favoring the modality with the best S/N on each trial. The noise level within a single modality would be more likely to be similar within a trial than noise levels in two different modalities. As a consequence, signal selection would be less effective in the dual single modality case. Note, however, that this strategy requires that subjects have some means of establishing which modality is more reliable on any given trial. It also requires that the independent fluctuations in S/N be sufficient to change which of the two modalities provides the more accurate signal estimate.

This selection theory is also compatible with the results of [Experiment 3](#), where no bimodal advantage was found. In this experiment, the most discriminable stimulus on each trial would be entirely determined by the added external noise.

## Conclusions

Discrimination, but not detection, of signal modulation rate was easier when stimuli were presented in two modalities rather than just one. This was true even when two unimodal signals were provided. Discrimination thresholds for repeated unimodal stimuli, in contrast, were not significantly different from the thresholds for single presentations of unimodal stimuli. The bimodal facilitation therefore did not require simultaneous presentations in both modalities and evidently does not rely on bimodal sensory fusion. We propose that the improved performance in bimodal over dual unimodal stimuli results from selection of the highest S/N stimuli within each trial. The noise level within a single modality is more likely to be similar within a trial than noise levels in two different modalities. As a consequence, signal selection would be more effective for bimodal stimulus presentations.

## Acknowledgments

This work was supported by the Human Frontier Science Program.

Commercial relationships: none.

Corresponding author: Ansgar Koene.

Email: arkoene@ntu.edu.tw.

Address: Psychology department, National Taiwan University, Taipei 106, Taiwan.

## Footnote

<sup>1</sup> According to the maximum likelihood estimation rule, the prediction for the discrimination threshold ( $T_{12}$ ) when

two cue stimuli are presented can be determined from the single stimulus discrimination thresholds ( $T_1$  and  $T_2$ ) as

$$T_{12}^2 = \frac{(T_1^2 T_2^2)}{(T_1^2 + T_2^2)}$$

(cf. Equation 7 of Ernst & Banks, 2002).

By design,  $T_1 = T_2$  in our experiment, regardless of whether  $T_1$  and  $T_2$  refer to the same stimulus modality (i.e., repeated unimodal conditions) or different modalities (i.e., bimodal conditions). Thus, we can write:

$$T_{12}^2 = \frac{(T_1^2 T_2^2)}{(T_1^2 + T_2^2)} = \frac{(T_1^2)^2}{(2T_1^2)} = \frac{T_1^2}{2} \Rightarrow T_{12} = \frac{T_1}{\sqrt{2}}.$$

## References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257–262. [[PubMed](#)] [[Article](#)]
- Andersen, T. S., Tiippana, K., & Sams, M. (2005). Maximum likelihood integration of rapid flashes and beeps. *Neuroscience Letters*, *380*, 155–160. [[PubMed](#)]
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A, Optics, image science, and vision*, *20*, 1391–1397. [[PubMed](#)]
- Bertelson, P., Vroomen, J., de Gelder, B., & Driver, J. (2000). The ventriloquist effect does not depend on the direction of deliberate visual attention. *Perception & Psychophysics*, *62*, 321–332. [[PubMed](#)]
- Bresciani, J. P., Ernst, M. O., Drewing, K., Bouyer, G., Maury, V., & Kheddar, A. (2005). Feeling what you hear: Auditory signals can modulate tactile tap perception. *Experimental Brain Research*, *162*, 172–180. [[PubMed](#)]
- Bresciani, J. P., Dammeier, F., & Ernst, M. O. (2006). Vision and touch are automatically integrated for the perception of sequences of events. *Journal of Vision*, *6*(5):2, 554–564, <http://journalofvision.org/6/5/2/>, doi:10.1167/6.5.2. [[PubMed](#)] [[Article](#)]
- Dayan, P., & Yu, A. J. (2002). Acetylcholine, uncertainty, and cortical inference. In T. G. Dietterich, S. Becker, & Z. Ghahramani (Eds.), *Advances in neural information processing systems* (vol. 14). Cambridge, MA: MIT Press.
- Ernst, M. O. (2006). A Bayesian view on multimodal cue integration. In G. Knoblich, I. M. Thornton, M. Grosjean, & M. Shiffrar (Eds.), *Human body*

- perception from the inside out*. New York: Oxford University Press.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433. [PubMed]
- Foster, D. H., & Bischof, W. F. (1987). Bootstrap variance estimators for the parameters of small-sample sensory-performance functions. *Biological Cybernetics*, *57*, 341–347. [PubMed]
- Hasselmo, M. E., & Fehrlau, B. P. (2001). Differences in time course of ACh and GABA modulation of excitatory synaptic potentials in slices of rat hippocampus. *Journal of Neurophysiology*, *86*, 1792–1802. [PubMed] [Article]
- Hillis, J. M., Ernst, M. O., Banks, M. S., & Landy, M. S. (2002). Combining sensory information: Mandatory fusion within, but not between, senses. *Science*, *298*, 1627–1630. [PubMed]
- Hillis, J. M., Watt, S. J., Landy, M. S., & Banks, M. S. (2004). Slant from texture and disparity cues: Optimal cue combination. *Journal of Vision* *4*(12):1, 967–992, <http://journalofvision.org/4/12/1/>, doi:10.1167/4.12.1. [PubMed] [Article]
- Kelly, D. H. (1979). Motion and vision. II. Stabilized spatio-temporal threshold surface. *Journal of the Optical Society of America*, *69*, 1340–1349. [PubMed]
- Landy, M. S., Maloney, L. T., Johnston, E. B., & Young, M. (1995). Measurement and modeling of depth cue combination: In defense of weak fusion. *Vision Research*, *35*, 389–412. [PubMed]
- Macmillan, N. A., & Creelman, C. D. (2005). *Detection theory: A user's guide* (2nd ed.). London: Lawrence Erlbaum Associates.
- Miller, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, *14*, 247–279. [PubMed]
- Patching, G. R., & Quinlan, P. T. (2004). Crossmodal integration of simple auditory and visual events. *Perception & Psychophysics*, *66*, 131–140. [PubMed] [Article]
- Pirenne, M. H. (1943). Binocular and unocular threshold of vision. *Nature*, *152*, 698–699.
- Quinlan, P. T., & Hill, N. I. (1999). Sequential effects in rudimentary auditory and visual tasks. *Perception & Psychophysics*, *61*, 375–384. [PubMed]
- Shams, L., Kamitani, Y., & Shimojo, S. (2000). Illusions. What you see is what you hear. *Nature*, *408*, 778. [PubMed]
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*, 147–152. [PubMed]
- Spence, C., & Driver, J. (1997). On measuring selective attention to an expected sensory modality. *Perception & Psychophysics*, *59*, 389–403. [PubMed]
- Stone, J. V., Hunkin, N. M., Porrill, J., Wood, R., Keeler, V., Beanland, et al. (2001). When is now? Perception of simultaneity. *Proceedings of the Royal Society B: Biological sciences*, *268*, 31–38. [PubMed] [Article]
- Treisman, M. (1998). Combining information: Probability summation and probability averaging in detection and discrimination. *Psychological Methods*, *2*, 252–265.
- Violentyev, A., Shimojo, S., & Shams, L. (2005). Touch-induced visual illusion. *Neuroreport*, *16*, 1107–1110. [PubMed]
- Vroomen, J., Bertelson, P., & de Gelder, B. (2001). The ventriloquist effect does not depend on the direction of automatic visual attention. *Perception & Psychophysics*, *63*, 651–659. [PubMed] [Article]
- Yuille, A. L., & Buelhoff, H. H. (1996). Bayesian decision theory and psychophysics. In D. C. Knill & W. Richards (Eds.), *Perception as Bayesian inference* (pp. 123–161). Cambridge University Press.