

# Multi-sensory integration of spatio-temporal segmentation cues: One plus one does not always equal two

Feng Zhou, Victoria Wong & Robert Sekuler  
Brandeis University, Waltham MA

How are multiple, multi-sensory stimuli combined for use in segmenting spatio-temporal events? For an answer, we measured the effect of various auditory or visual stimuli, in isolation or in combination, on a bistable percept of visual motion (“bouncing” vs. “streaming”). To minimize individual differences, the physical properties of stimuli were adjusted to reflect individual subjects’ sensitivity to each cue in isolation. When put into combination, perceptual influences that had been equipotent in isolation were substantially altered. Specifically, auditory cues that had been strong when presented alone, were greatly reduced in combination. Evaluation of alternative models of sensory integration showed that the state of the visual bistable percept could not be accounted for by probability summation among cues, as might occur at the level of decision processes. Instead, the state of the bistable percept was well predicted from a weighted sum of cues, with visual cues strongly dominating auditory cues. Finally, when cue weights were compared for individual subjects, it was found that subjects differ somewhat in the strategy they use for integrating multi-sensory information.

## Introduction

Making sense of the perceptual world requires that the spatio-temporal stream of sensory information be appropriately segmented into constituent parts. By distinguishing sensory inputs that arise from distinct events, or from distinct components of complex actions, successful segmentation promotes percepts that are meaningful, temporally stable, and behaviorally-useful. With a moving object, abrupt changes in the object’s speed or direction of movement are essential data for segmentation, either by human observers (Zacks et al., 2001), or by non-biological, event-parsing systems (Rui & Anandan, 2000; Agam, Bullock, & Sekuler, 2005). Under many conditions, visual segmentation cues are accompanied by correlated inputs from other senses, notably audition (Roskies, 1999; Shimojo & Shams, 2001). Content-based, automated image recognition systems can improve their segmentation of events by exploiting correlated multi-sensory information (Rui, Gupta, & Acero, 2000), but can humans do the same? And if they can, by what rules are multiple segmentation cues integrated? To answer these questions we devised a novel paradigm to evaluate individual and joint effects of visual and auditory cues on the bistable percept of a motion stimulus.

Seven decades ago, Metzger (1934) introduced a family of bistable stimuli whose members are well-suited for studying segmentation. From viewers’ perceptions of the paths taken by objects moving along intersecting trajectories, Metzger identified conditions that promoted perceptual segmentation. Such conditions caused an object moving along a single continuous trajectory to appear to move along first one and then another, different trajectory. Recently, several researchers have implemented Metzger’s stimuli as two identical discs that move smoothly across a computer display on opposed, intersecting paths (Bertenthal, Banton, & Bradbury, 1993; Sekuler & Sekuler, 1999; Watanabe & Shimojo, 2001a). The discs’ momentary coincidence on the display sets up a perceptual ambiguity: the discs can either appear to pass through one another (“streaming”), or appear to collide and bounce off one another (“bouncing”). Some added cues, such as a brief pause in the discs’ motion, support segmentation of either disc’s movement into two distinct trajectories and promote “bouncing.”

Because the resulting state of this bistable “bouncing-streaming” percept is sensitive to auditory as well as visual cues, the percept is a good vehicle for exploring multi-sensory contributions to segmentation. Sekuler, Sekuler, and Lau (1997) showed that a sound added to the bistable visual stimulus biased observers to perceive “bouncing” motion. This demonstration of multi-sensory integration has been replicated and extended several times in psychophysical studies (for example, Remijn, Ito, & Nakajima, 2004; Sanabria, Correa, Lupianez, & Spence, 2004; Scheier, Lewkowicz, & Shimojo, 2003; Watanabe & Shimojo, 2001b; Sakurai & Grove, 2006). Moreover, in a study combining psychophysics and functional neuroimaging, Bushara et al. (2003) found significant differences in the patterns of brain activation associated with the two perceptual re-

---

We thank Larry Abbott, Yuko Yotsumoto, Takeo Watanabe, Allison B. Sekuler, and Kristina Visscher for excellent suggestions. Feng Zhou is now at the Department of Psychological and Brain Sciences, the Johns Hopkins University. Victoria Wong was supported by an NSF IGERT fellowship, and by an Undergraduate Research grant; she is currently at the School of Medicine, University of Hawaii. Research supported by AFOSR grant F49620-03-1-0376 and National Institutes of Health grant MH-55687. e-mail: sekuler@brandeis.edu

sponses. Notably, trials on which the sound successfully promoted “bouncing” showed strong interactions between unimodal and multimodal sensory areas, with the unimodal activations associated with “streaming” diminished, and unique multimodal area activation emerging. Although these results are insufficient to support quantitative descriptions of interactions between unimodal and multimodal areas of the brain, they do suggest that auditory and visual influences on the bistable stimulus might not sum linearly (see also, Lewis, Beauchamp, & DeYoe, 2000).

The present study sought to quantify the ways in which human observers combined multi-sensory cues that are added to the bistable motion stimulus. For example, we asked whether the combination of multi-sensory cues could be fully explained by probability summation on a decision level after cue components had been processed independently of one another (Wuerger, Hofbauer, & Meyer, 2003; Alais & Burr, 2004a), or if combination of multi-sensory cues reflects integration at a sensory level. The latter form of integration could encompass modality specific as well as cross-modal sensory processing as seen in superior colliculus neurons of cats (eg. Stanford, Quessy, & Stein, 2005; Meredith & Stein, 1986), and in human psychophysical results (Meyer, Wuerger, Röhrbein, & Zetzsche, 2005).

Recent work has generated quantitative descriptions of how multi-sensory cues are integrated at a sensory level. For example, Ernst and Banks (2002) offered an elegant description of how human observers reconciled discrepant visual and haptic cues to yield a final estimate of an object’s size. In their task, observers assigned weights to individual sensory inputs in proportion to their *a priori* reliability. More specifically, the weight for a sensory cue was the normalized reciprocal variance of the cue. Sensory integration in their task was thus well described by a weighted mean of the separate influences, with the sum of weights clamped to a constant value of unity. In contrast, Meyer et al. (2005)’s study of motion detection showed that auditory and visual motion signals sum linearly, at least when they were co-localized. Finally, combining psychophysics and functional neuroimaging, Beauchamp (2005) demonstrated that with combined, multi-sensory inputs, the size of human observers’ Blood Oxygenation Level Dependent (BOLD) signals fell between the mean of responses to the individual uni-sensory inputs and the sum of those responses.

The diversity of previous findings led us to examine multi-sensory integration of cues that could be used to aid spatio-temporal segmentation. We began by evaluating various visual and auditory cues’ efficacy in causing discs that moved along intersecting paths to appear to “bounce” rather than “stream.” Then we measured the combined effects of pairs or trios of cues, tailoring the physical value of each cue to suit individual observers’ sensitivity to that cue. With diverse sensory cues transformed into a common metric, we next compared each cue’s perceptual effectiveness when presented in isolation to its effectiveness when presented in combination with other cues. Finally, we evaluated a suite of quantitative models representing alternative accounts of cue-cue interaction. To anticipate, for nearly all subjects,

the Akaike Information Criterion (Akaike, 1973) identified as best a model in which the state of the bistable percept is governed by a weighted sum of cues, but with individual cues differing substantially in their individual influence upon the whole.

## Method

To examine their impact on spatio-temporal segmentation, we added various cues, singly or in combination, to a perceptually-bistable, moving stimulus. Extensive pilot testing showed that cues differed in potency, and that individuals differed somewhat in their sensitivity to various cues. Therefore, we adjusted the cues parametrically so that (i) all cues presented individually would be equally effective, and (ii) cues were scaled to reflect individual differences among observers. We then determined how well the cues’ equivalence was preserved when they were combined, making quantitative comparisons of the perceptual result of these combinations to predictions from a suite of alternative models.

### Stimuli

*Bi-stable Motion Stimulus.* At the start of each trial, a 0.5 deg black fixation cross appeared in the center of a computer display maintained at a uniform gray color of 41.86 cd/m<sup>2</sup> luminance. Immediately after the fixation cross appeared, two black discs appeared, 6.7° to either side of the fixation cross. After 1-sec, the fixation cross disappeared and the discs immediately began moving at 5.93 deg/sec. Each disc moved steadily toward the position originally occupied by its mate; they coincided at the center of the display. Then, after reaching their destinations, the discs disappeared. The discs were each 1.09° in diameter, and had a luminance of 1.08 cd/m<sup>2</sup>. At the end of a trial, the subject pressed one of two keys, indicating whether the discs had appeared to bounce off one another (“bouncing”) or seemed to stream through one another (“streaming”).

Stimuli were presented on a 15” cathode ray display, which was refreshed by an Apple iMac computer at 96 Hz. Each subject viewed the computer display binocularly, head steadied by a forehead rest and chin cup 57 cm from the display. Stimuli were formatted as QuickTime movies, generated in Macromedia Flash MX, and presented under the control of Showtime (Watson & Hu, 1999), a component of the Psychophysics Toolbox (Brainard, 1997; Pelli, 1997) extensions to Matlab. Showtime’s frame by frame control of display rate made it possible to produce the small variations that were needed to quantify some cues’ effects.

*Multi-sensory Cues.* Sekuler and Sekuler (1999) showed that alterations in some characteristics of a visual bistable stimulus changed perception, promoting a percept of bouncing. For example, introducing a brief pause or an occluding surface into the display increased the proportion of “bouncing” judgments. Sekuler et al. (1997) demonstrated an inter-modal influence on the perception of the bistable visual stimulus, showing that, under some conditions, the addition of a sound could promote perceived bouncing. To the same

bistable stimulus we added four cues, two auditory and two visual, each of which was meant to act as a segmentation signal, conveying information that the discs' paths had been perturbed. In preliminary experiments, each of these added cues successfully promoted a percept of bouncing. Before selecting these four cues, we explored various other cues that we thought might influence perception of the bistable motion stimulus. Some cues, such as shape deformation, which accompanies collisions between non-rigid objects, had no discernible effect on perceived bouncing. Of the other cues tested, four were selected for their strong, reliable influence on perceived bouncing. These four cues, which were used in the actual experiments, were: Visual Luminance (*VL*): Luminance of discs at the moment of their overlap; Visual Duration (*VD*): Duration of pause at overlap; Auditory Intensity (*AI*): Intensity of a click sound at overlap; and Auditory Timing (*AT*): Sound onset time relative to overlap.

For visual cues, we varied the duration (*VD*) for which the coincident discs overlapped, or we momentarily reduced the discs' contrast by increasing the black discs' luminance (*VL*) for the single frame during which the discs were coincident. In condition *VD*, the duration of the pause during the display's middle frame varied over the range from 25 to 150 msec. In condition *VL*, the relative luminance of the discs at the middle frame (point of coincidence) varied randomly from trial to trial, with disk luminances ranging from 1.15 to 12.07 cd/m<sup>2</sup>, relative to the 1.08 cd/m<sup>2</sup> normal luminance of discs, and the 41.86 cd/m<sup>2</sup> luminance of the background. As the discs' momentary contrast decreased,  $P(\text{bounce})$  increased.

For both auditory cues, a single tapping sound was added to the bistable stimulus, which otherwise was silent. The spectrum of the 35-msec sound peaked near 2 kHz. The sound was delivered via a pair of speakers (Yamaha YSTM15), which had a reasonably flat frequency response over the range 70 Hz to 20 kHz ( $\pm 3$  dB(A)). The speakers were located symmetrically on either side of the computer display to mimic real environment in which visual and sound information of the same event tended to co-localize. We varied either the timing of the tap's presentation, or its sound pressure level. To manipulate Auditory Timing (*AT*), we varied the interval between the tap's onset and the display frame in which the discs were coincident. For this cue, the tap's sound pressure level was constant at 85 dB(A). The time at which the tap was presented was varied from 0 to 576 msec before the discs' coincidence point. To manipulate Auditory Intensity (*AI*), we varied the tap's sound pressure level, keeping its timing constant, that is, it was always presented at the start of the frame during which the discs overlapped. For *AI*, the tap's intensity varied over a narrow range, from 47.5 to 49.5 dB(A), which was sufficient to generate the full span of psychometric response.

### Procedure

To generate psychometric functions for the four different cues, each was added individually to the bistable stimulus. To span the psychometric range from  $0 < P(\text{bounce}) < 1$ , be-

tween 8 and 16 different levels were used for each of the four cues. Each subject made "streaming-bouncing" judgments on a minimum of 20-24 trials for each cue type and level.

Individual psychometric functions were used to estimate, for each subject and cue type, stimulus cue values that corresponded to one low level ( $\sim 20\%$ ) and one higher level ( $\sim 30\%$ ) of  $P(\text{bounce})$ . To avoid hitting the upper limit of  $P(\text{bounce}) = 1$  in the case that three cues linearly summed their effects, we set both the low-level and the higher-level cue intensities relatively weak, that is,  $P(\text{bounce}) < 33\%$ . The two levels of cue intensity provided two equipotent sets of four cue types for each individual, allowing cross-modal combination at different stimulus intensity levels. Any low or high level cue was combined with both low and high levels of any other cues, in pairs or in trios. Because the psychometric results with *AT* was generated with one fixed value of *AI*, these two auditory cues, *AT* and *AI*, were not combined parametrically as other cues were. The combined multisensory cues were added to the baseline bistable stimulus. For each subject, the 36 combinations of two and three cues were presented 24 times each, in random order.

### Subjects

Subjects were ten Brandeis University undergraduates who were naive to the purpose of the study. Of these subjects, two were dropped because their psychometric functions for one or more individual cues produced indeterminate estimates of criterion values for those cues. The indeterminacy made it impossible to identify the cue values required for further testing. One remaining subject's data were excluded because his performance was utterly uncorrelated with that of the other participants, mean Pearson's  $r = -0.05$  ( $SD = 0.004$ ). Written consent was obtained from the participants and they were made aware that their involvement was voluntary and could be withdrawn at any time. The experimental procedures were approved by Brandeis University's Institutional Review Board.

## Results

### Individual Cues

Figure 1 shows one subject's psychometric functions that resulted when each of the individual cues was added to the bistable stimulus. The various functions indicate that subjects generally perceived the basic visual motion stimulus as "streaming" when no additional cue was added, or when a weak cue was presented in the same session with stronger cues. This result suggests that "streaming" is a kind of default percept for this basic motion stimulus, and as various cues were added, the resulting percept (either "streaming" and "bouncing") became bistable. The smooth curve in each panel represents the subject's best fitting Weibull function, determined by maximum likelihood. The Weibull cumulative probability function was defined as:

$$P(\text{bounce}) = 1 - e^{-\left(\frac{x}{\alpha}\right)^\beta} \quad (1)$$

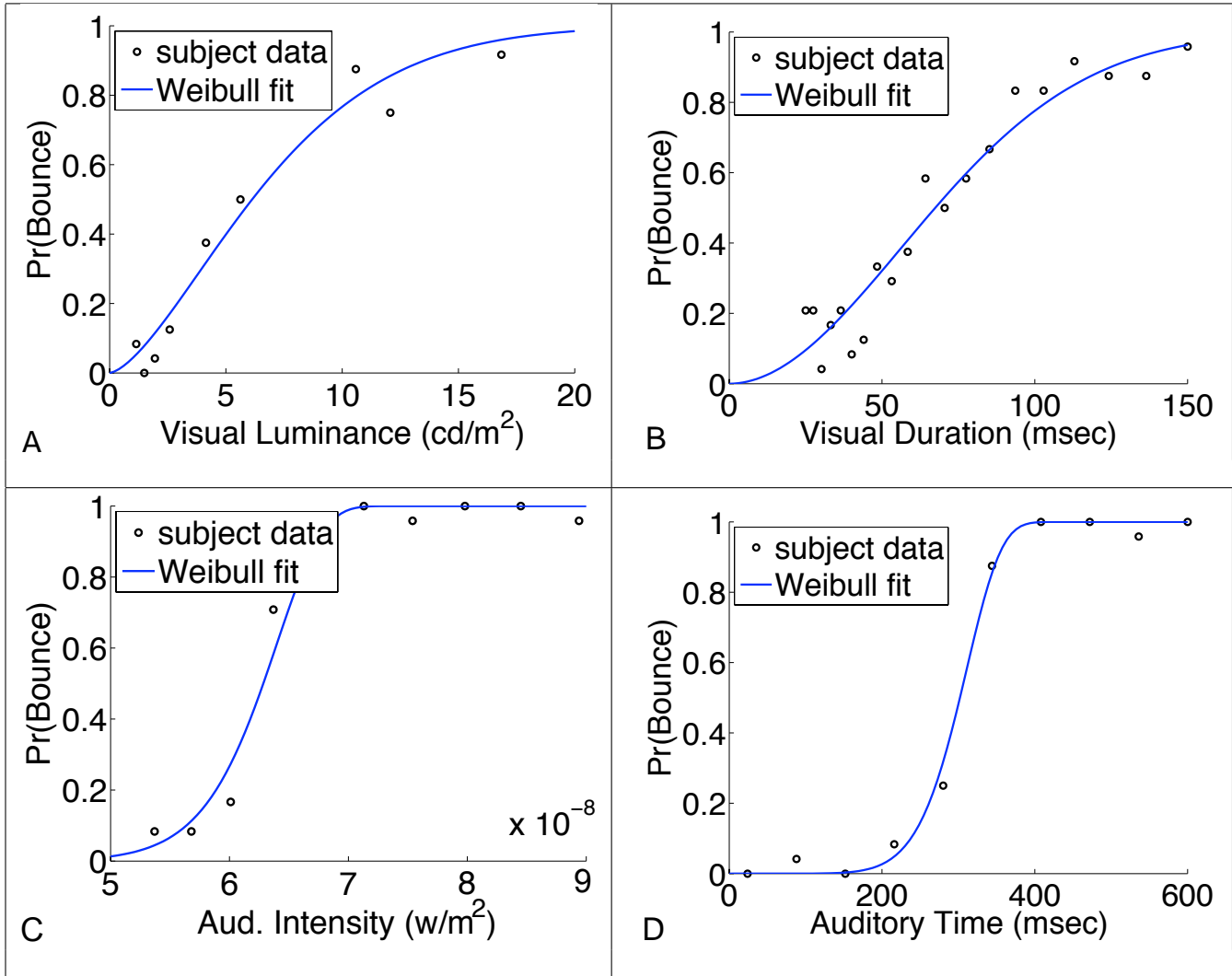


Figure 1. Psychometric functions produced when each of the four cues is added to the ambiguous display. Curves represent the best fitting Weibull function. Panel A: fit for cue VL ( $\alpha=7.8$ ,  $\beta=1.51$ ); Panel B: fit for cue VD ( $\alpha=86.2$ ,  $\beta=1.93$ ); Panel C: fit for cue AI ( $\alpha=6.42e-8$ ,  $\beta=17.34$ ); Panel D: fit for cue AT ( $\alpha=316$ ,  $\beta=7.87$ ). Data are for Subject 8.

Equation 1 describes probability of perceiving “bounce” as a function of stimulus intensity  $x$ . In this Weibull function,  $P(\text{bounce})$  increases from 0 to 1 as  $x$  varies over a sufficient range, and the function takes a sigmoidal shape, with a floor at 0 and a ceiling at 1. The Weibull function reflects Weber’s law such that if a signal increases logarithmically, its output increases in a way similar to a cumulative normal distribution. Thus on a logarithmic  $x$ -axis, changing the  $\alpha$  value linearly displaces the Weibull function horizontally, while changing the  $\beta$  value adjusts the function’s slope. Furthermore, because the function always crosses the point of  $(\alpha, 0.63)$ , intensity at level  $\alpha$  always produces 63% “bounce” response (Treutwein, 1995). Because of these properties, the Weibull cumulative function is frequently used to define psychometric functions, including binary distributions of two-alternative-force-choice responses such as the psychometric

functions in our experiment (Mortensen, 2002).

We tested alternative functions that allowed non-zero lapse and guess rates, but because the simpler function fit the data as well as the more complex functions did, subsequent analyses used the simpler version in Equation 1.

Figure 1A shows sample results for cue VL. As luminance of the moving discs was momentarily increased,  $P(\text{bounce})$  increased. Note that the normal luminance of the discs was  $1.08 \text{ cd/m}^2$  (left end of horizontal axis), and that as disc luminance increased toward the steady background value of  $41.86 \text{ cd/m}^2$ , the overlapping discs became increasingly less distinct. However, even the highest disc luminance used was still less than half of the background luminance, and the discs were clearly visible. Figure 1B shows results for condition VD. As the duration of disc overlap lengthened,  $P(\text{bounce})$  increased. Figure 1C shows that as the tap sound intensity

(AI) increased, so too did  $P(\text{bounce})$ . Finally, Figure 1D shows that as the presentation of the tap sound grew closer to the time of the discs coincidence,  $P(\text{bounce})$  increases. Because the Weibull function requires  $x \geq 0$ , we defined the coincident time to be 600 ms, which set the earliest sound onset in the experiment to be 24 ms. Table 1 summarizes  $\alpha$  and  $\beta$  parameter values of the Weibull functions for individual subjects.

Table 1  
*Weibull Function Parameters for Individual Subjects*

Subj	Visual Luminance		Visual Duration	
	$\alpha$	$\beta$	$\alpha$	$\beta$
1	4.007	2.916	74.503	4.053
2	6.826	1.254	80.163	3.127
4	2.312	5.107	81.81	3.265
5	6.177	1.495	74.909	1.801
6	3.534	4.356	60.03	4.962
7	6.645	0.965	142.423	1.007
8	7.808	1.516	81.353	1.95
<i>M</i>	5.329	2.515	85.027	2.881
Subj	Auditory Intensity		Auditory Time	
	$\alpha$	$\beta$	$\alpha$	$\beta$
1	9.852	7.274	694.939	8.745
2	6.329	14.265	449.971	3.484
4	6.136	24.691	447.027	3.655
5	7.427	4.411	434.588	4.379
6	6.843	13.006	425.821	12.942
7	7.147	5.959	386.599	2.273
8	6.419	17.339	316.418	7.872
<i>M</i>	7.165	12.421	450.766	6.193

### Normalizing Individual Cues

The cues used in the experiment covered different ranges and were expressed in incommensurable units (e.g.,  $\text{cd}/\text{m}^2$ , ms, and  $\text{w}/\text{m}^2$ ). So prior to investigating how multi-sensory cues were combined, we needed some way to translate the four separate cues into a common metric. For this purpose, we projected various cue intensities onto a common Weibull function template:

$$P(\text{bounce}) = 1 - e^{-\left(\frac{z}{10}\right)^{3.5}} \quad (2)$$

where  $z$  is the standardized stimulus value of one cue, either presented alone or in combination with other cues. We arbitrarily chose  $\alpha=10$  so that when  $z=10$ ,  $P(\text{bounce}) \approx .63$ . We chose  $\beta=3.5$  because the psychometric functions for many sensory responses are well fitted by a Weibull function with  $\beta = 3.5$  (Watson & Pelli, 1983).

Rearranging Equation 2 yields:

$$z = 10 \times (-\ln(1 - P))^{1/3.5} \quad (3)$$

We first calculated  $P(\text{bounce})$  according to individual subjects' Weibull functions for cue intensity. Then we inverted

probability to standardized intensity,  $z$ , using Equation 3. The resulting standardized intensity of various cues were now commensurable and unit independent. And regardless of its raw value, a larger standardized intensity caused more "bouncing" percept than a smaller one. Different combining rules were then applied to these standardized intensities to generate a composite input,  $z'$ . These composite inputs, reflecting various combining rules, yielded predictions for  $P(\text{bounce})$ . These predictions were used to evaluate alternative models of multisensory integration.

### Formal Accounts of Multi-sensory Integration

We began our investigation of how cues are combined in our task by examining the relation between  $P(\text{bounce})$  and the number of cues present in a multi-cue stimulus. Figure 2 shows the mean  $P(\text{bounce})$  produced by various single cues, and by 2-cue and 3-cue combinations, uni-modal as well as multi-modal. Note that on average, the two levels of intensities in single-cue conditions produced about 20% and 30%  $P(\text{bounce})$ , confirming that our equalizing procedure indeed produced equivalent efficacy across four different cues. Furthermore, the presence of three cues promotes a higher proportion of bouncing than the presence of two cues ( $p < .05$ ). However, by itself the number of cues present in the combination does not determine  $P(\text{bounce})$ ; in fact, the set of 2-cue combinations comprises a substantial range of  $P(\text{bounce})$  values, as does the set of 3-cue combinations. The integrated effects of cues clearly depend on how they were combined rather than on their effects in isolation. Specific combinations and the corresponding  $P(\text{bounce})$  values are listed in a table in supplementary materials.

To characterize cue interactions quantitatively, we cast the multi-cue results into a suite of alternative frameworks, each representing a different description of the interaction among cues. These alternative frameworks, which are described below, included one in which responses to cues were processed independently of one another, with any empirical signs of interaction being merely a reflection of probability summation (Sanabria, Spence, & Soto-Faraco, 2006). The other three frameworks represented the proposition that cues did interact, either with all cues having equal effects, or with cues differing in their effects. The following paragraphs give brief descriptions of each of the four alternative theoretical frameworks.

- **Probability Summation Model.** Influences from multiple cues might be combined at a decision level rather than at a perceptual one. In some studies of visual and auditory detection, observers seemed to integrate multiple cues via probability summation (for example, Meyer et al., 2005; Wuerger et al., 2003). Assuming that the integration of multiple cues occurs at a decision level, and that  $P(\text{bounce})$  values for various cues in combination are independent of one another, we can use a standard probability summation formula to predict values of  $P(\text{bounce})$  when  $n$  sensory inputs are combined:

$$P(\text{bounce}) = 1 - \prod_{i=1}^n (1 - P_i) \quad (4)$$

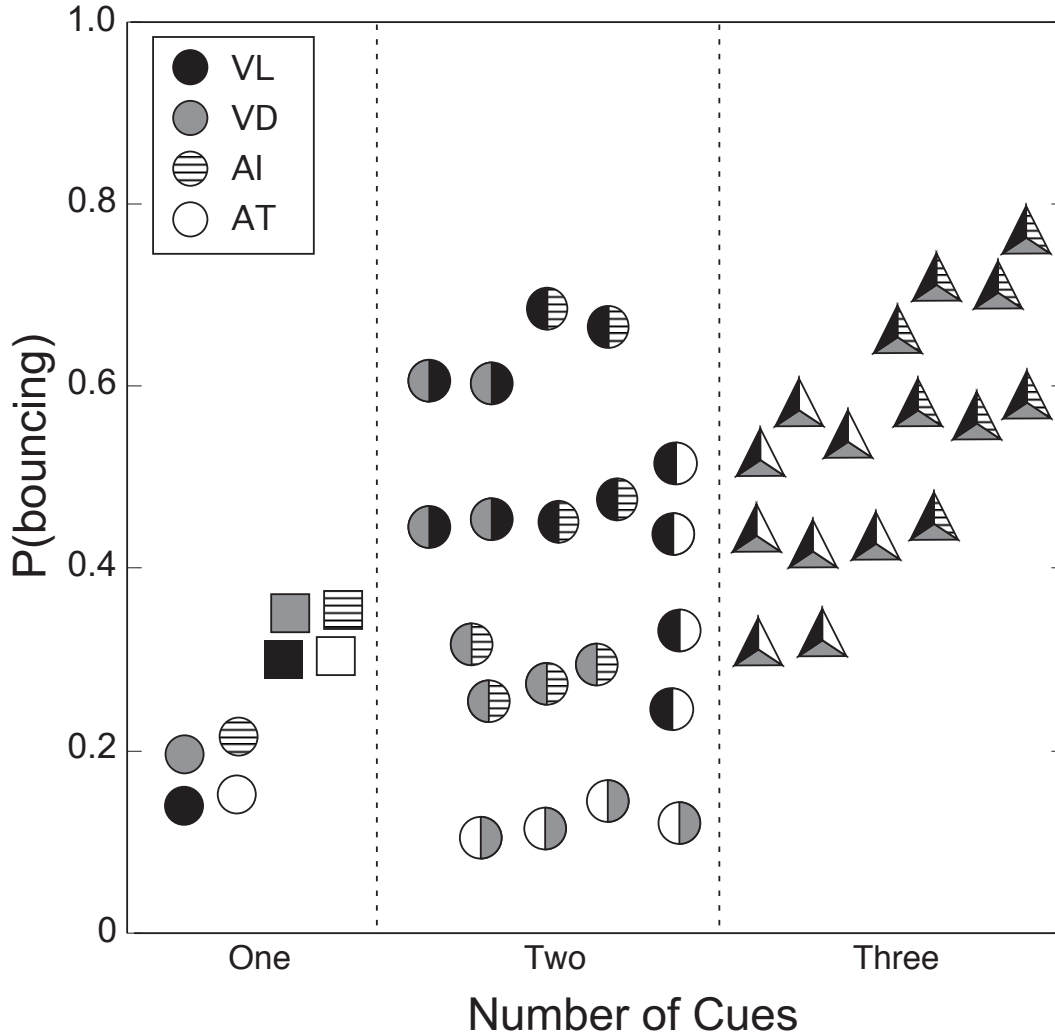


Figure 2. Mean  $P(\text{bounce})$  associated with single cues, and with two-cue and and three-cue combinations. Conditions involving VL, VD, AI and AT are shown by black, gray, striped, and white symbols, respectively. For single-cue conditions, lower levels of each cue are represented as circles; the higher level of each cue are represented by squares. For the sake of legibility, the symbols do not distinguish between low and higher levels for two- and three-cue combinations.

where  $P_i$  is the probability of a bouncing response for the  $i^{\text{th}}$  sensory input when it is presented alone.

- **Summation Models.** The remaining three alternative accounts assume that cues' effects are not independent of one another, but are integrated at a sensory level. Each cue effect,  $z_i$ , is integrated into a quantity  $z'$ , which is the sensory basis for the decision. The probability of “bouncing” response from the integrated input  $z'$  can be calculated from Equation 2. We examined three variants of this general hypothesis.

*Equal Weight Summation.* This hypothesis asserts that cues sum with equal weights to produce an integrated sensory signal on which the “bouncing” vs. “streaming” response is based. With any combination of cues, their standardized values are weighted equally. Namely, individual cues' effects in combination are proportional to their initial effects in isolation.

$$z' = \theta \sum_{i=1}^n (z_i) \quad (5)$$

where  $z_i$  is the  $i^{\text{th}}$  intensity of  $n$  sensory inputs, standardized according to Equation 3;  $\theta$  is a constant weight for all senses when they are combined, and  $z'$  is the integrated input that affects perception. Note that an individual cue's effect is not necessarily the same as the effect of that cue when presented in isolation; depending upon the value of  $\theta$ , all cues' contribution in combination may be scaled up or scaled down by  $\theta$ .

*Variable Weight Summation.* This hypothesis asserts that cue effects sum, but do so with weights that may vary among cues. That is, cues that are equally effective when presented singly, may not contribute equally to the sum when they are combined. This simple alternative, which does not

allow for non-linear interactions among cue effects, can be described as

$$z' = \sum_{i=1}^n (\theta_i z_i) \quad (6)$$

where  $\theta_i$  represents the weight assigned to the  $i^{\text{th}}$  cue.

*Variable Weights & Interactions* This hypothesis asserts that cue effects sum, but do so with weights that can vary among cues, and that cue effects may interact. This account specifically allows for the possibility that the effect of one cue can vary, depending on the intensity levels of other cues.

$$z' = \sum_{i=1}^n (\theta_i z_i) + \sum_{j=1}^m (\gamma_j \cdot \text{interaction}_j) \quad (7)$$

In this model,  $\gamma_j$  is a weight assigned to the  $j^{\text{th}}$  interaction term. Each non-linear interaction is modeled as the product of standardized intensities of two or more cues. Our experiment's design yields  $m = 7$  interaction terms:  $VL \times VD$ ,  $VL \times AI$ ,  $VL \times AT$ ,  $VD \times AI$ ,  $VD \times AT$ ,  $VL \times VD \times AI$ , and  $VL \times VD \times AT$ .

For purposes of model selection, the four alternative accounts—probability summation and the three variants of sensory summation—comprise “nested models” (Myung, Pitt, & Kim, 2005; Myung & Pitt, 2002). We exploited this nesting relationship in evaluating each model's fit to individual subjects' results from conditions with multiple cues. A downhill simplex search algorithm (Nelder & Mead, 1965) optimized the parameter values for each model by maximizing the likelihood of that model given an empirical data set. The likelihood was calculated as:

$$\mathcal{L}(\text{model}|\text{data}) = \prod_{i=1}^n \left[ \frac{(a_i + b_i)!}{a_i! b_i!} p_i^{a_i} (1 - p_i)^{b_i} \right] \quad (8)$$

where  $n$  is the total number of conditions (number of data points in Figure 2);  $a_i$  and  $b_i$  were the number of observed “bouncing” and “streaming” trials comprising the  $i^{\text{th}}$  condition;  $p$  is the  $pr(\text{bounce})$  predicted by the model.

The resulting fits were evaluated using the Akaike Information Criterion ( $AIC$ ) and each model's Akaike weights,  $w(AIC)$ . In order to simultaneously consider the generalizability and flexibility of these models to select a better model (Myung & Pitt, 2002), we calculated  $AIC$  for each model:

$$AIC = -2 \log_e \mathcal{L} + 2k \quad (9)$$

where  $\mathcal{L}$  is the likelihood of a model given a data set as defined in Equation 8, and  $k$  is the number of free parameters in a model. A more complex model in a series of nested models can always provide equally good or better fit than simpler ones even if the extra parameters were spurious, because the more complex model enjoys extra degrees of freedom.  $AIC$  penalizes the more complex model to the extent that the gain in goodness of fit is more than offset by the cost of spurious parameters. Because  $\mathcal{L}$  is a small fractional in Equation 9, it can be seen that  $AIC$  is a positive number often in the order hundreds, and a smaller value of  $AIC$  indicates a better model.

$AIC$  by itself only provides a binary interpretation of a comparison. Namely, a smaller  $AIC$  indicates a more likely model, but it does not specify how much better it is. Wagenmakers and Farrell (2004) developed an  $AIC$  weight method to measure the comparative likelihood of multiple models on a continuous scale. We calculated  $AIC$  Weights as:

$$w_i(AIC) = \frac{e^{-\frac{1}{2}\Delta_i(AIC)}}{\sum_{j=1}^m e^{-\frac{1}{2}\Delta_j(AIC)}} \quad (10)$$

and

$$\Delta_i(AIC) = AIC_i - \min(AIC) \quad (11)$$

where  $w_i(AIC)$  is the  $AIC$  Weight for the  $i^{\text{th}}$  model,  $\min(AIC)$  is the minimal  $AIC$  among the  $m$  models being compared. Ranging from 0 and 1,  $AIC$  Weight indicates the probability that a particular model is the “best” among a set of compared models.

Table 2 shows the optimized models' error, expressed as root mean squared difference ( $RMSD$ ), and the associated  $AIC$  weights. For every subject but number six, the  $AIC$  weights associated with the four models show that the *Variable Weights* model is best. Comparing the *Variable Weight* summation model with its *Variable Weight Summation & Interaction* counterpart, the  $RMSD$  values show that the more complex model, with interaction parameters, fits the data slightly better. However, except for Subject 6, the slight improvement of fit is insufficient to offset the the penalty assessed for having additional parameters.

Figure 3A-G shows the *Variable Weight* model's fit to each subject's data. Also shown are  $r^2$  and  $RMSD$ , which reflect goodness of fit of that model to the data. Note that some subjects'  $pr(\text{bounce})$  does not evenly span the whole range between 0 and 1. For instance, subject 2 has a cluster of  $pr(\text{bounce})$  near 0.1. These data points correspond to conditions in which one weak visual cue is combined with an auditory cue. As can be seen in Table 3, subject 2 assigned very small weights to auditory cues. The combination of a weak visual cue and an auditory cue was therefore not effective in causing a bouncing percept, giving rise to a cluster in the lower left corner of subject 2's data panel. Similarly, the data points in the upper right corner correspond to conditions combining two strong visual cues. The clustering of data in Figure 3 reflects the fact that for some subjects certain cues weight very little, while other cues weight much more heavily.

Having found that the *Variable Weights* model provides the best account of the multiple-cue data, we examined the values of the individual weights that optimized the model's fit. Table 3 displays the optimized weights for individual subjects. Note that for five of the seven subjects,  $VD$  has the largest weight. And for all subjects but one,  $AI$  and  $AT$  have the smallest weights. So, generally, the two visual weights,  $VD$  and  $VI$ , boast the strongest influence on the bistable percept, with the two auditory weights,  $AI$  and  $AT$ , having much less influence. A repeated-measure ANOVA confirmed that the weights differed significantly from one another,  $F(3,18) = 13.464$ ,  $p < .0001$ . Prominent among

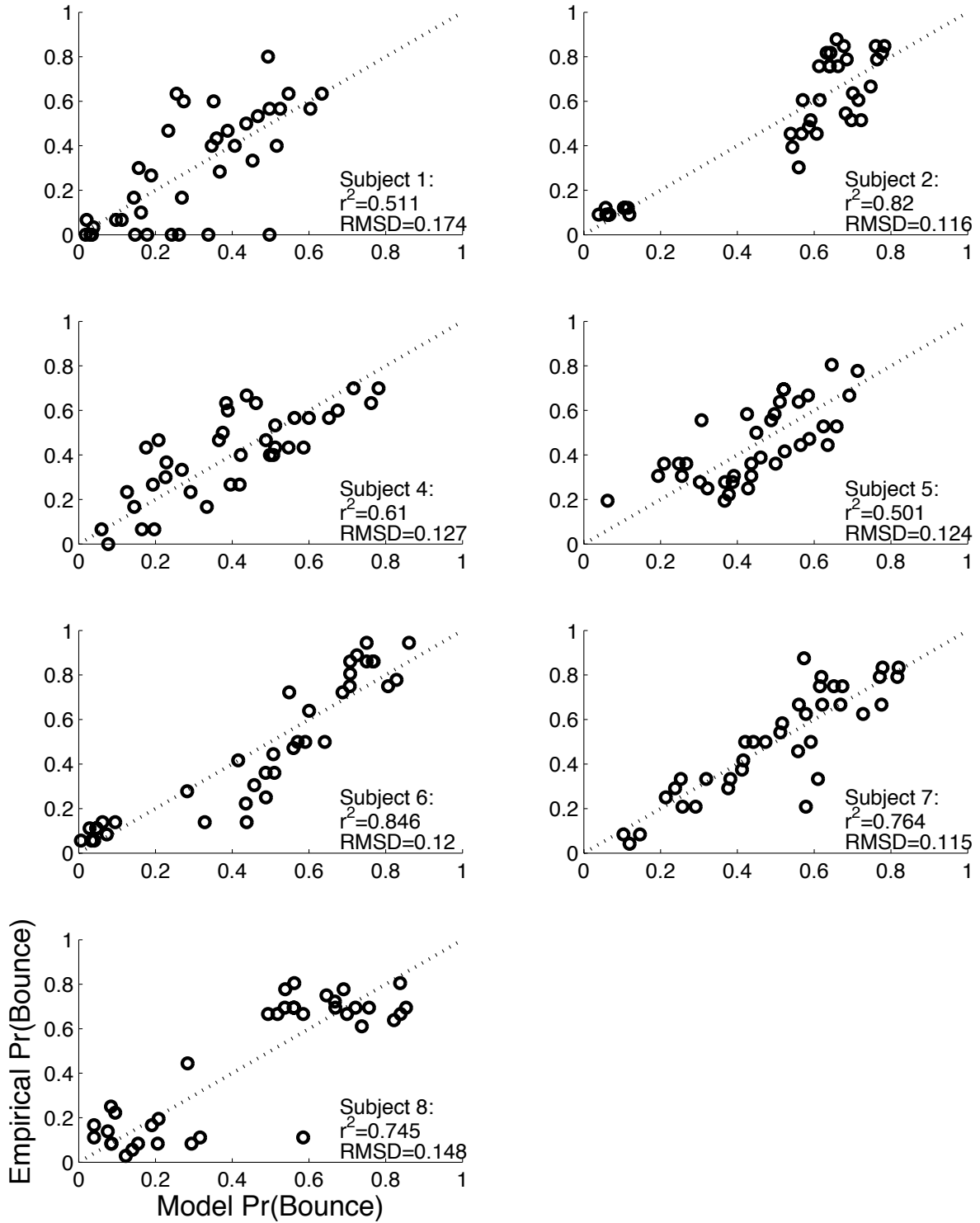


Figure 3. Observed  $P(\text{bounce})$  versus  $P(\text{bounce})$  as predicted by the Variable Weights model. Each panel represents results for one subject, and gives the associated values of  $r^2$  and root mean square error (RMSD).

Table 2  
*RMSD and  $w(AIC)$  for Each Model by Subject*

Subj	Alternative models							
	Probability Summation		Equal Weight Summation		Variable Weight Summation		Variable Weight Sum. & Interaction	
	<i>RMSD</i>	$w(AIC)$	<i>RMSD</i>	$w(AIC)$	<i>RMSD</i>	$w(AIC)$	<i>RMSD</i>	$w(AIC)$
1	0.285	0	0.214	0	0.174	0.825¶	0.164	0.175
2	0.242	0	0.205	0	0.116	0.996¶	0.111	0.004
4	0.202	0	0.214	0	0.127	0.998¶	0.127	0.002
5	0.199	0	0.140	0	0.124	0.999¶	0.124	0.001
6	0.267	0	0.207	0	0.12	0.001	0.092	0.999¶
7	0.191	0	0.214	0	0.115	0.998¶	0.113	0.002
8	0.324	0	0.360	0	0.148	0.99 ¶	0.146	0.01

¶ The subject's most likely model.

the relationships among weights was substantial reduction in auditory cues' effects when they were combined with visual cues. A pairwise comparison between the weaker visual cue *VL* and the stronger auditory cue *AI*, shows that, too, the visual cue contributes significantly more weight than the auditory cue ( $p < .05$ ). In fact, for the weaker auditory cue, *AT*, the weight did not significantly differ from zero,  $t(6) = -0.82$ ,  $p > 0.44$ . Note that for four subjects *AT* took on a negative value, showing that the auditory cue exerted an inhibitory effect as the audiovisual asynchrony was reduced. This inhibitory effect probably reflects some higher-order interactions between the *AT* cue and visual cues. In interpreting this substantial divergence between the modalities' effects in multi-cue combinations, it must be borne in mind that all cues, regardless of modality, had been equated for their individual impacts according to their strengths presented in isolation, as evident in the tightly clustered values of single-cue conditions in Figure 2.

Table 3  
*Weights for Variable Weight Summation Model by Subject*

Subject	Cue			
	<i>VL</i>	<i>VD</i>	<i>AI</i>	<i>AT</i>
1	0.564	0.634	0.172	-0.129
2	0.815	0.643	0.040	0.022
4	0.369	1.038	0.078	-0.245
5	0.293	0.770	0.257	0.189
6	1.028	0.346	0.189	0.279
7	0.423	0.998	0.053	-0.177
8	0.132	1.208	0.187	-0.543
<i>M</i>	0.518	0.808	0.139	-0.090
<i>SEM</i>	0.118	0.110	0.031	0.105

### Non-linear Interactions Among Weights

As mentioned earlier, the models of cue interaction evaluated here were nested, that is, the number of free parameters increased across models, reaching a maximum in the

model that incorporated variable weights as well as interactions. As a model selection instrument, the Akaike Information Criterion penalizes a model for including additional parameters. The  $w(AIC)$  values in Table 2 show that, taking the parameter penalty into account, for only one subject does the  $w(AIC)$  favor the model that includes Variable Weights and Interactions, over the model that excludes interactions. Although the  $w(AIC)$  values gave a decisive nod to the *Variable Weights* model that does not include interactions, it is still useful to examine the pattern of interactions that were turned up when interactions were included.

Note first the individual cues' weights similarity here (leftmost four columns of data) to the corresponding weights for the *Variable Weights* model from which interactions were excluded (see Table 3). This consistency reflects the very modest values that were generally assumed by the interaction terms in the more complex model. It also supports the idea that non-linear interactions among cues indeed play a very modest role in the process of multi-sensory integration.

The last seven columns of Table 4 show the values of the five two-way and the two three-way interaction parameters in the optimized *Variable Weight & Interaction* model. Earlier, we noted that just one subject's  $w(AIC)$  favored the model that included interactions over the comparable model with none. The reason for this can be seen in Table 4's interaction values for that subject, Subject 6. Nearly every interaction weight for that subject is substantial; the one exception being the weight for the interaction between the two visual cues, *VL* and *VD*. Note also that Subject 6's model weight for cue *VD* is 0, which deviates greatly from the high weight accruing to that cue with all other subjects.

At the other extreme from Subject 6 are Subjects 4 and 5 whose results provide the most clear-cut demonstration that the model's interaction terms are genuinely inconsequential. The other subjects fall between these extremes.

Table 4  
*Variable Weight & Interaction Model Parameters by Subject*

Subj	Cue Weights				Interactions							
	<i>VL</i>	<i>VD</i>	<i>AI</i>	<i>AT</i>	<i>VL</i>	<i>VL</i>	<i>VL</i>	<i>VD</i>	<i>VD</i>	<i>VL</i>	<i>VL</i>	
					<i>VD</i>	<i>AI</i>	<i>AT</i>	<i>AI</i>	<i>AT</i>	<i>VD</i>	<i>VD</i>	
1	0.14	0.72	0.38	-0.63	0.29	-0.39	0.56	-0.30	-0.04	0.52	0.04	
2	0.96	0.48	0.16	-0.09	0.06	-0.13	-0.05	0.02	0.34	-0.06	-0.24	
4	0.40	1.06	0.15	-0.35	-0.06	-0.02	-0.01	-0.06	0.01	0.01	0.12	
5	0.29	0.77	0.25	0.20	0.01	0.02	-0.01	-0.01	0.01	0.01	-0.02	
6	1.53	0.00	0.86	-0.61	-0.02	0.65	0.26	-0.17	1.32	-1.25	-0.72	
7	0.55	0.84	0.09	-0.20	0.06	-0.08	-0.19	0.12	0.22	-0.12	-0.04	
8	0.20	1.18	0.11	-0.34	-0.02	0.02	-0.06	0.18	-0.26	-0.15	0.14	
<i>M</i>	0.58	0.72	0.28	-0.29	0.04	0.01	0.07	-0.03	0.23	-0.15	-0.10	

## Discussion

### *Sensory Integration With Variable Weights*

We identified visual and auditory cues that effectively influenced spatial-temporal segmentation of a bistable motion stimulus. In this study, we combined multiple cues from both visual and auditory modalities; previous investigations have included only one cue from each modality (e.g., Heron, Whitaker, & McGraw, 2004; Roach, Heron, & McGraw, 2006; Alais & Burr, 2004b). Furthermore, unlike many other studies of sensory integration (eg. Ernst & Banks, 2002; Soto-Faraco, Spence, Lloyd, & Kingstone, 2004), information of multiple cues in our study were not discrepant or conflicting. Our empirical and modeling results show that when these multi-sensory cues are presented jointly, their integration is not a simple probability summation as might be carried out on a purely decision level (Wuerger et al., 2003). Instead, it is likely a perceptual integration process much like summation process described in the *Variable Weights* model, with equally-effective unisensory cues substantially changing their effectiveness when placed into combination.

### *Dominance of Visual Cues*

When placed in combination, visual cues exerted much stronger influence on the perceptual state of the bistable percept than did auditory cues. Roach et al. (2006) showed that when sensitivity to auditory and visual information were equalized, both senses can affect each other to an equal extent. In contrast, the visual dominance in our task held even when all cues had been matched for their uni-sensory effectiveness. The two auditory cues, *AI* and *AT*, which in isolation were effective in inducing a “bouncing” percept, lost nearly all their influence when they were combined with visual cues. Of course, vision’s dominance over audition has been observed in other settings, including studies of visual capture (Posner, Nissen, & Klein, 1976; Colavita, 1974) and ventriloquism of motion (Soto-Faraco, Lyons, Gazzaniga, Spence, & Kingstone, 2002; Soto-Faraco et al., 2004). And there are situations in which the reverse relationship is reported, with auditory information dominating visual, as in

auditory capture (Shams, Kamitani, & Shimojo, 2002). Additionally, one cue’s complete dominance over others has been observed when multiple cues were all from the same modality (Bülhoff & Mallot, 1988). We can consider some candidate explanations for visual cues’ dominance over their auditory counterparts in our paradigm.

First, a Bayesian approach to sensory integration implies that if cues differ in reliability, more reliable cues are advantaged over less reliable ones (Ernst & Banks, 2002; Knill & Saunders, 2003; Alais & Burr, 2004b). The most obvious available index of reliability is the slope of the psychometric function associated with any cue. In particular, a steeper slope signals that a cue is more reliable. Because not all of our cues are expressed in the same physical units, comparisons of slopes for some cue pairs are beyond reach. As a result, it is not possible to make all the comparisons that could bear on predictions from a Bayesian perspective. However, there is one cross-modal pair in which both visual and auditory cues are expressed in the same units. These cues are *VL* and *AT*, which are both expressed in units of msec. Table 1’s  $\beta$  values reveal that on average, the psychometric functions for *VD* were less steep (mean = 2.881) than those for *AT* (mean=6.193). This difference is confirmed by a pair-wise *t*-test, which produced  $t(6)=2.96$ ,  $p < .05$ . So, in a manner not consistent with a Bayesian framework, the cue that seems to be less precise is weighted more heavily in combination with a cue that is more precise.

Second, however, this straightforward extrapolation from variation in the slopes of psychometric functions may fail to take account of an important task dependency. Extending the basic Bayesian approach, Welch and Warren (1986) proposed that discrepancies between senses tend to be resolved in favor of the modality that is not only generally more precise, but was also more appropriate to the task at hand. For example, in temporal judgements, audition tends to be the more appropriate modality and therefore tends to dominate, but in spatial judgments, vision tends to prevail (for example, see Witten & Knudsen, 2005). Although stimulus motion, like that represented in our task, generates both spatial and temporal information, motion characteristics might be most

precisely estimated using *visual* information rather than *auditory* information. This advantage would make vision the more appropriate source of information for motion-related tasks, and thus the default choice between vision and audition. Reviewing the literature on auditory motion perception, Middlebrooks and Green (1991) concluded that auditory motion perception was nothing more than detection of changes in static location over time, and saw no evidence for the kind of motion-sensitive mechanisms known to support the perception of visual motion.

Some recent studies support this second idea that for multisensory integration in motion perception, vision is the more appropriate modality. In one study, Berryhill, Chiu, and Hughes (2006) compared the gain of smooth pursuit eye movements for visual and various non-visual motion stimuli. The gain of an smooth pursuit eye movement is defined as the ratio of eye velocity to stimulus velocity. The researchers found an average gain of just under 0.7 with visual motion, but only 0.1 with auditory motion. Of all forms of motion tested, including tactile and proprioceptive motion, gain was by far the lowest for audition. These findings are generally consistent with the interference relationships among vision, audition and touch reported by Soto-Faraco et al. (2004). In another study, Soto-Faraco, Spence, and Kingstone (2004) found that when auditory and visual apparent motion streams were presented concurrently in opposite directions, participants often failed to discriminate the auditory motion direction, whereas perception of the visual motion was unaffected by the direction of auditory motion, and this asymmetry persisted even when the perceived quality of apparent motion was equated for the two modalities. Heron et al. (2004) further combined the modality appropriateness idea with Bayesian approach, suggesting that vision is the default modality for motion perception, and that the use of the default depends on the certainty of available information. Specifically, they propose that when visual information is certain, vision is the default modality for motion perception, but it gives way to audition as the uncertainty associated with visual information rises relative to the uncertainty associated with audition. Although these previous results do not lead directly to a precise, quantitative prediction about the degree of cue dominance that should be expected when visual and auditory cues are combined, they are consistent with the direction of dominance that is observed in our study, converging on the notion that auditory information seems to be neither precise nor appropriate in estimating motion.

### *Sensory Reconciliation and Sensory Combination*

The strategies that sensory systems use to integrate information vary with the nature of the stimulus information and the task. We distinguish between two broad classes of integration strategies, namely “sensory reconciliation” and “sensory summation”. In this dichotomy, “sensory reconciliation” covers cases in which information is conflicting, redundant and commensurable. A major part of multisensory integration literature addresses such situations (for reviews, see De Gelder & Bertelson, 2003; Ernst & Bühlhoff, 2004). In

contrast, “sensory combination” strategy is used when information is neither redundant nor commensurable. Our study is one of the few that investigate the integration strategy in such a task. We believe that these two categories provide a good framework within which to compare processes that are involved in various integration tasks.

The *Variable Weight Summation Model* describes the integrated sensory signal as a weighted sum of the individual sensory inputs (Equation 6). This principle was also used in recent studies of “sensory reconciliation” strategy (Ernst, Banks, & Bühlhoff, 2000; Ernst & Banks, 2002; Hillis, Ernst, Banks, & Landy, 2002). In those studies, subjects estimated the surface properties of an object when visual and haptic information was discrepant. Human observers tended to resolve discrepancies between cues by resorting to an intermediate value, that is, a compromise. Under those conditions, weights of visual and auditory cues were inversely related, with weights summing to a constant value:

$$\sum_{i=1}^n \theta_i = 1$$

In contrast to judgments of surface properties, the temporal-spatial segmentation that underlies our task seems to require a “sensory combination” strategy. According to Gestalt principles, the trajectories in our task tend to be seen as temporally continuous and spatially smooth (Tripathy & Barrett, 2003). This perceptual default condition would cause each moving disc in our bistable stimulus to appear to move uninterrupted through the other disc (Sekuler & Sekuler, 1999). A shift away from this default state might require sufficient perceptual evidence that a disc’s motion comprised not one, but two distinct temporal episodes, one preceding the discs’ coincidence, and one following it (Zacks & Tversky, 2001). Cues that signal a perturbation in the discs’ movement likely promote path segmentation and therefore the percept of “bouncing”. In combining signals that arise from multiple cues in our task, subjects seem to use a “sensory combination” strategy, in which the sum of weights is not clamped at some constant value:

$$\sum_{i=1}^n \theta_i > 1$$

A special case of “sensory combination” involves super-additivity of multiple sensory inputs, in which observers maximize non-redundant information, such as in motion detection with weak, near-threshold sensory signals (eg. Meyer et al., 2005; Wallace, Meredith, & Stein, 1998). In this case, the sum of weights can grow with the number of cues:

$$\sum_{i=1}^n \theta_i \geq n$$

As noted above, various non-redundant multi-sensory cues in “sensory combination” may be incommensurable. The novel approach we used to standardize cues made possible quantitative comparisons of the influence exerted by various cues. This procedure can potentially be applied to other

stimuli and tasks to further investigate how weights are assigned to various cues in “sensory combination”.

### *Non-linear Interactions and Individual Differences*

Most models of sensory combination assume that any interactions among the effects of various cues are additive. Although that assumption may be appropriate in treating the results of some or even many subjects in our task, results from the *Variable Weight & Interactions* model suggest that subjects may actually fall along a continuum with respect to the extent of non-linear interactions among cues. At one extreme would be Subjects 4 and 5, whose results showed near-zero interactions, and at the other extreme would be Subject 6, whose results showed substantial interactions among cues. Similar variability of cue interaction was also observed in a visuotactile perception study (Guest & Spence, 2003), in which different subjects’ texture discrimination performance were best described by different models: averaging, linear summation, as well as power law summation models. Future research clearly needs to assess the stability of such individual differences in our task, a test that would require far more subjects than we have studied, as well as appropriately spaced retests of those subjects. If individual differences prove to be stable, one might want to know whether such individual differences also make significant contributions to the patterns of integration seen in other multi-sensory cue integration tasks. Finally, one would certainly want to understand the origin of these individual differences, including the possible involvements of top-down influences such as memory (Lewkowicz, 1996; Scheier et al., 2003) and attention (Ernst & Bühlhoff, 2004; McDonald, Teder-Salejarvi, DiRusso, & Hillyard, 2005).

### References

- Agam, Y., Bullock, D., & Sekuler, R. (2005). Imitating unfamiliar sequences of connected linear motions. *Journal of Neurophysiology*, *94*, 2832-2843.
- Akaike, H. (1973). A new look at statistical model identification. *IEEE Transactions of Automatic Control*, *19*, 716-723.
- Alais, D., & Burr, D. (2004a). No direction-specific bimodal facilitation for audiovisual motion detection. *Cognitive Brain Research*, *19*, 185-94.
- Alais, D., & Burr, D. (2004b). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*, 257-62.
- Beauchamp, M. (2005). Statistical criteria in fMRI studies of multisensory integration. *Neuroinformatics*, *3*, 93-113.
- Berryhill, M. E., Chiu, T., & Hughes, H. C. (2006). Smooth pursuit of nonvisual motion. *Journal of Neurophysiology*, *96*, 461-465.
- Bertenthal, B. I., Banton, T., & Bradbury, A. (1993). Directional bias in the perception of translating patterns. *Perception*, *22*, 193-207.
- Brainard, D. H. (1997). The psychophysics toolbox. *Spatial Vision*, *10*, 433-436.
- Bühlhoff, H. H., & Mallot, H. A. (1988). Integration of depth modules: stereo and shading. *Journal of the Optical Society of America, A*, *5*, 1749-1758.
- Bushara, K., Hanakawa, T., Immisch, I., Toma, K., Kansaku, K., & Hallett, M. (2003). Neural correlates of cross-modal binding. *Nature Neuroscience*, *6*, 190-5.
- Colavita, F. B. (1974). Human sensory dominance. *Perception & Psychophysics*, *16*, 409-412.
- De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*, 460-467.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429-33.
- Ernst, M. O., Banks, M. S., & Bühlhoff, H. H. (2000). Touch can change visual slant perception. *Nature Neuroscience*, *3*, 69-73.
- Ernst, M. O., & Bühlhoff, H. H. (2004). Merging the senses into a robust percept. *Trends in Cognitive Sciences*, *8*, 162-169.
- Guest, S., & Spence, C. (2003). What role does multisensory integration play in the visuotactile perception of texture? *International Journal of Psychophysiology*, *50*(-), 63-80.
- Heron, J., Whitaker, D., & McGraw, P. (2004). Sensory uncertainty governs the extent of audio-visual interaction. *Vision Research*, *44*, 2875-84.
- Hillis, J., Ernst, M., Banks, M., & Landy, M. (2002). Combining sensory information: mandatory fusion within, but not between, senses. *Science*, *298*, 1627-30.
- Knill, D., & Saunders, J. (2003). Do humans optimally integrate stereo and texture information for judgments of surface slant? *Vision Research*, *43*, 2539-58.
- Lewis, J. W., Beauchamp, M. S., & DeYoe, E. A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. *Cerebral Cortex*, *10*, 873-88.
- Lewkowicz, D. J. (1996). Perception of auditory-visual temporal synchrony in human infants. *Journal of Experimental Psychology: Human Perception & Performance*, *22*, 1094-106.
- McDonald, J. J., Teder-Salejarvi, W. A., DiRusso, F., & Hillyard, S. A. (2005). Neural basis of auditory-induced shifts in visual time-order perception. *Nature Neuroscience*, *8*, 1197-1202.
- Meredith, M. A., & Stein, B. E. (1986). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, *56*, 640-662.
- Metzger, W. (1934). Beobachtungen über phänomenale Identität (observations on phenomenal identity). *Psychologische Forschung*, *19*, 1-60.
- Meyer, G., Wuerger, S., Röhrbein, F., & Zetzsche, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*.
- Middlebrooks, J. C., & Green, D. M. (1991). Sound localization by human listeners. *Annual Review of Psychology*, *42*, 135-159.
- Mortensen, U. (2002). Additive noise, weibull functions and the approximation of psychometric functions. *Vision Res*, *42*, 2371-93.
- Myung, I. J., & Pitt, M. A. (2002). Mathematical modeling. In J. Wixted (Ed.), *Stevens' handbook of experimental psychology* (3 ed., p. 429-459). Wiley.
- Myung, J., Pitt, M. A., & Kim, W. (2005). Model evaluation, testing and selection. In K. Lambert & R. Goldstone (Eds.), *Handbook of cognition*. Sage Publication.
- Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *Computing Journal*, *7*, 308-313.

- Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*, 437-442.
- Posner, M., Nissen, M., & Klein, R. (1976). Visual dominance: an information-processing account of its origins and significance. *Psychological Review*, *83*, 157-71.
- Remijn, G. B., Ito, H., & Nakajima, Y. (2004). Audiovisual integration: An investigation of the "streaming-bouncing" phenomenon. *Journal of Physiological Anthropology: Applied Human Sciences*, *23*, 243-247.
- Roach, N., Heron, J., & McGraw, P. (2006). Resolving multisensory conflict: a strategy for balancing the costs and benefits of audiovisual integration. *Proceedings of the Royal Society Biology*, *273*, 2159-68.
- Roskies, A. L. (1999). The binding problem. *Neuron*, *24*, 7-9.
- Rui, Y., & Anandan, P. (2000). Segmenting visual actions based on spatio-temporal motion patterns. In *IEEE Proceedings of Conference on Computer Vision and Pattern Recognition* (p. 1111-1118). Hilton Head, South Carolina.
- Rui, Y., Gupta, A., & Acero, A. (2000). Automatically extracting highlights for TV baseball programs. In *Proceedings of ACM Multimedia* (p. 105-115). Los Angeles.
- Sakurai, K., & Grove, P. M. (2006). Auditory induced bounce perception when visual trajectories are inconsistent with motion reversal. In *European Conference on Visual Perception*. St. Petersburg, Russia.
- Sanabria, D., Correa, A., Lupianez, J., & Spence, C. (2004). Bouncing or streaming? Exploring the influence of auditory cues on the interpretation of ambiguous visual motion. *Experimental Brain Research*, *157*, 537-541.
- Sanabria, D., Spence, C., & Soto-Faraco, S. (2006). Perceptual and decisional contributions to audiovisual interactions in the perception of apparent motion: A signal detection study. *Cognition*.
- Scheier, C., Lewkowicz, D. J., & Shimojo, S. (2003). Sound induces perceptual reorganization of an ambiguous motion display in human infants. *Developmental Sciences*, *6*, 233-241.
- Sekuler, A. B., & Sekuler, R. (1999). Collisions between moving visual targets: what controls alternative ways of seeing an ambiguous display? *Perception*, *28*, 415-432.
- Sekuler, R., Sekuler, A. B., & Lau, R. (1997). Sound alters visual motion perception. *Nature*, *385*, 308.
- Shams, L., Kamitani, Y., & Shimojo, S. (2002). Visual illusion induced by sound. *Cognitive Brain Research*, *14*, 147-152.
- Shimojo, S., & Shams, L. (2001). Sensory modalities are not separate modalities: plasticity and interactions. *Current Opinion in Neurobiology*, *11*, 505-509.
- Soto-Faraco, S., Lyons, J., Gazzaniga, M., Spence, C., & Kingstone, A. (2002). The ventriloquist in motion: illusory capture of dynamic information across sensory modalities. *Cognitive Brain Research*, *14*, 139-46.
- Soto-Faraco, S., Spence, C., & Kingstone, A. (2004). Cross-modal dynamic capture: congruency effects in the perception of motion across sensory modalities. *Journal of Experimental Psychology: Human Perception and Performance*, *30*, 330-45.
- Soto-Faraco, S., Spence, C., Lloyd, D., & Kingstone, A. (2004). Moving multisensory research along: Motion perception across sensory modalities. *Current Directions in Psychological Science*, *13*, 29-32.
- Stanford, T., Quessy, S., & Stein, B. (2005). Evaluating the operations underlying multisensory integration in the cat superior colliculus. *Journal of Neuroscience*, *25*, 6499-508.
- Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, *35*, 2503-22.
- Tripathy, S. P., & Barrett, B. T. (2003). Gross misperceptions in the perceived trajectories of moving dots. *Perception*, *32*, 1403-1408.
- Wagenmakers, E.-J., & Farrell, S. (2004). AIC model selection using Akaike weights. *Psychonomic Bulletin & Review*, *11*, 192-196.
- Wallace, M., Meredith, M., & Stein, B. (1998). Multisensory integration in the superior colliculus of the alert cat. *Journal of Neurophysiology*, *80*, 1006-10.
- Watanabe, K., & Shimojo, S. (2001a). Postcoincidence trajectory duration affects motion event perception. *Perception & Psychophysics*, *63*, 16-28.
- Watanabe, K., & Shimojo, S. (2001b). When sound affects vision: Effects of auditory grouping on visual motion perception. *Psychological Science*, *12*, 109-116.
- Watson, A., & Hu, J. (1999). Showtime: A Quicktime-based infrastructure for vision research displays. *Perception (Supplement)*, *28*, 45.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113-120.
- Welch, R., & Warren, D. (1986). Intersensory interactions. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 25.1 - 25.36). New York City: Wiley.
- Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron*, *48*, 489-496.
- Wuerger, S., Hofbauer, M., & Meyer, G. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics*, *65*, 1188-96.
- Zacks, J. M., Braver, T. S., Sheridan, M. A., Donaldson, D. I., Snyder, A. Z., Ollinger, J. M., Buckner, R. L., & Raichle, M. E. (2001). Human brain activity time-locked to perceptual event boundaries. *Nature Neuroscience*, *4*, 651-655.
- Zacks, J. M., & Tversky, B. (2001). Event structure in perception and conception. *Psychological Bulletin*, *127*, 3-21.