Article
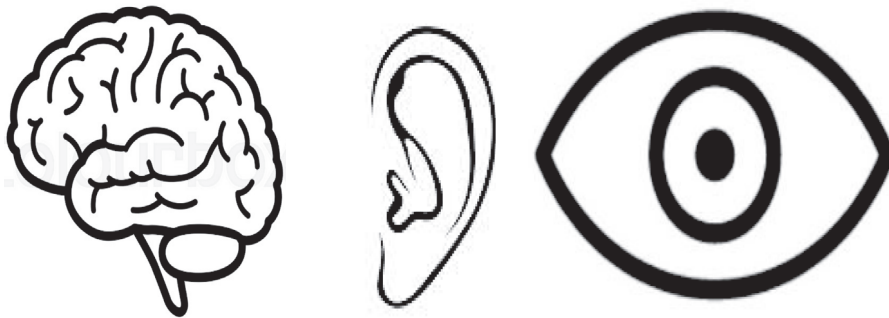
A simple vector-like law for perceptual information combination is also followed by a class of cortical multisensory bimodal neurons

Audiovisual =

$$(\text{Audio}^{1.7} + \text{Visual}^{1.7})^{1/1.7}$$

Vincent A. Billock, Micah J. Kinney, Jan W.H. Schnupp, M. Alex Meredith

vincent.billock.ctr@us.af.mil

Highlights

The neural correlates of perceptual information combination rules are unknown

Little is known about precisely how bimodal neurons combine two sensory signals

Both perceptual and cortical data exhibit vector-like audiovisual combination

Cortical bimodal neurons may underlie audiovisual sensitivity

Article

A simple vector-like law for perceptual information combination is also followed by a class of cortical multisensory bimodal neurons

Vincent A. Billock,[1,5,*] Micah J. Kinney,[1,2] Jan W.H. Schnupp,[3] and M. Alex Meredith[4]

SUMMARY

An interdisciplinary approach to sensory information combination shows a correspondence between perceptual and neural measures of nonlinear multisensory integration. In psychophysics, sensory information combinations are often characterized by the Minkowski formula, but the neural substrates of many psychophysical multisensory interactions are unknown. We show that audiovisual interactions – for both psychophysical detection threshold data and cortical bimodal neurons – obey similar vector-like Minkowski models, suggesting that cortical bimodal neurons could underlie multisensory perceptual sensitivity. An alternative Bayesian model is not a good predictor of cortical bimodal response. In contrast to cortex, audiovisual data from superior colliculus resembles the 'City-Block' combination rule used in perceptual similarity metrics. Previous work found a simple power law amplification rule is followed for perceptual appearance measures and by cortical subthreshold multisensory neurons. The two most studied neural cell classes in cortical multisensory interactions may provide neural substrates for two important perceptual modes: appearance-based and performance-based perception.

INTRODUCTION

Many models of sensory information combinations use the Minkowski equation (Equation 1), a modified Pythagorean sum created by an eminent pioneer of non-Euclidean geometry and relativity theory (Minkowski, 1891, 1910).

$$SensorySummation\ (A, B, \ldots Z) = (Channel_A{}^m + Channel_B{}^m + \ldots + Channel_Z{}^m)^{1/m} \quad \text{(Equation 1)}$$

Various sensory information combinations are associated with m values ranging from 1 (the 'City Block' model sometimes used in perceptual similarity studies (Attneave, 1950; Coombs et al., 1970; Garner, 1974; Landahl, 1945; Wuerger et al., 1995) to about 8 (Quick (1974) famous probability summation model). As m increases, summation becomes increasingly nonlinear, with more weight placed on the strongest sensory input (Coombs et al., 1970). Although some studies associate various values of m with neural theory-based interpretations (Landahl, 1945; Lehky, 1983; To et al., 2010), the most attention has been paid to the significance of $m = 2$; vector-like models and their associated Euclidean-like metric properties are important in experimental psychology (Shepard, 1964, 1987; Coombs et al., 1970; Garner, 1974; To et al., 2010). Roger Shepard argued strongly for vector-like information combinations throughout sensory and cognitive science (Shepard, 1987) and vector models have been of particular importance in color vision (Guth and Lodge, 1973; Ingling and Tsou, 1977; Guth et al., 1980). Curiously, there has been little effort to understand the neural correlates of vector-like sensory combinations. However, recent work in multisensory integration shows promise for understanding the neural correlates of some vector-like sensory combinations. In particular, one of us has found evidence for vector-like summation in multisensory detection thresholds (Schnupp et al., 2005). Figure 1A shows data for one observer that has a best fit m of 1.78. When we reanalyzed Schnupp et al. (2005) data, we found that for all three observers in three lighting conditions, the average m was 1.67 ± 0.26 (sd). This raises an interesting opportunity because there is an extensive database on neural multisensory interactions that can be examined for correspondence to the perceptual data.

Of the known mechanisms, the most plausible neural correlates for multisensory vector-like summation are facilitatory cortical bimodal neurons. In superior colliculus/optic tectum and in sensory cortex, there are

[1]Naval Aerospace Medical Research Laboratory, NAMRU-D, Wright-Patterson Air Force Base, OH 45433, USA

[2]Naval Air Warfare Center, NAWCAD, Patuxent River, MD 20670, USA

[3]Department of Neuroscience, City University of Hong Kong, Kowloon Tong, Hong Kong, China

[4]Department of Anatomy and Neurobiology, Virginia Commonwealth University, Richmond, VA 23298, USA

[5]Lead contact

*Correspondence: vincent.billock.ctr@us.af.mil

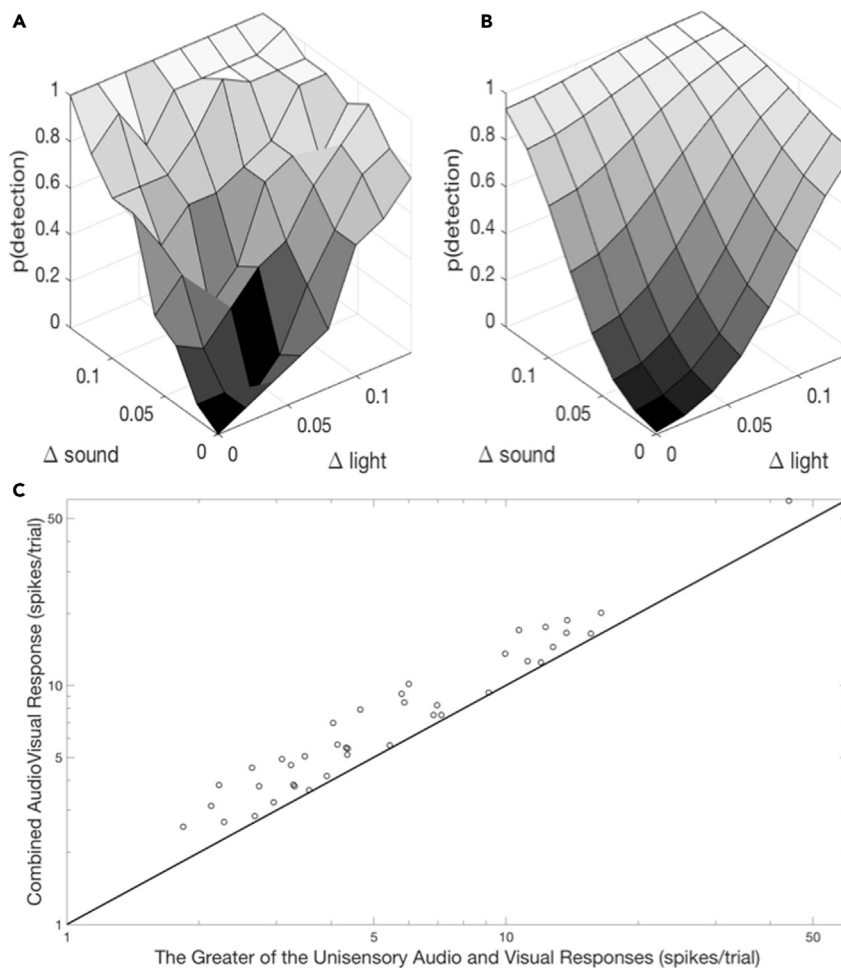https://doi.org/10.1016/j.isci. 2021.102527

Figure 1. A disconnect between perceptual and neural approaches for understanding multisensory integration

(A) Representative perceptual data from Schnupp et al. (2005), showing proportion of successful detections of a change of an audiovisual stimulus as a function of the amplitude of the change (Weber fraction). A Minkowski model fit to this data yields a Minkowski exponent m of 1.78.

(B) A vector model ($m = 2$) fit to the data shown in (A).

(C) Data from 40 facilitatory PLLS bimodal neurons in cat visual cortex (Meredith et al., 2012). Although both studies investigate responses to bimodal (audio and visual) stimuli, the analysis of bimodal neural data almost always discards data about the weaker unisensory response. The standard neural data representation plots the multisensory response as a function of the best unisensory response. This is useful for illustrating multisensory facilitation (the height of a multisensory response above the plotted line is a measure of facilitation), but it gives no insight into multisensory combination rules (see Figure 2 for the same neural data set, but with the discarded weaker sensory responses restored in a modeling-based framework).

bimodal neurons which nonlinearly combine sensory signals from various modalities. For example, in rattlesnake optic tectum there are neurons that are driven by visual signals from the eyes and by infrared signals from the heat sensitive facial pits that give pit vipers their name (Hartline et al., 1978; Newman and Hartline, 1981). Although these bimodal cells will fire to either visible light or heat, they fire harder when pits and eyes are simultaneously stimulated (Newman and Hartline, 1981). Similar cells have been extensively studied in rabbits, owls, guinea pigs, cats, ferrets and primates, for responses to combinations of audio, visual and tactile information (Horn and Hill, 1966; King and Palmer, 1985; Meredith & Stein, 1983, 1985, 1986; Stein and Meredith, 1993; Zahar et al., 2009). Bimodal neurons are thought to combine information from separate modalities nonlinearly, and the strength of the combination varies. In cat, for example, sensory combinations in bimodal cortical neurons generally yield modest multisensory enhancements (Meredith et al., 2012). In contrast, bimodal neurons in cat superior colliculus can show extraordinary multisensory responses (Meredith and Stein, 1986; Stein and Meredith, 1993).

Bimodal neurons are thought to play a role in some perceptual multisensory combinations (Stein et al., 1988), yet no one has attempted to analyze bimodal data and perceptual multisensory data in a unified information combination framework (Figure 1 shows the missing connection between perceptual and neural studies). This is not a simple oversight – the crux of the problem is that neural studies do not directly address multisensory combination but rather quantify multisensory enhancement (the relationship between the stronger unisensory response and the combined multisensory response). This approach disregards the weaker unisensory contribution. If the missing weaker unisensory responses for the multisensory neural data are included, then neural (Meredith et al., 2012) and perceptual (Schnupp et al., 2005) data – derived from our combined laboratories – can be analyzed in a common multidisciplinary framework developed for this purpose. Here we analyze neural data from two cortical areas that have many audiovisual bimodal neurons: Area Anterior Ectosylvian Sulcus (AES) (multisensory association cortex) and Area Posterolateral Suprasylvian Sulcus (PLLS), an extrastriate visual cortex in Brodmann area 19 (similar to Area MT in primates) that borders auditory cortex. We model data from cortical audiovisual bimodal neurons with the Minkowski equation, and we show that these neurons behave similarly to a vector-like (m = 2) model. More importantly, Minkowski exponents derived from the cortical audiovisual bimodal neurons closely resemble the Minkowski exponents for audiovisual perceptual threshold detection data (Schnupp et al., 2005), suggesting that cortical bimodal neurons could underlie audiovisual perceptual sensitivity. Because superior colliculus neurons are known for their remarkably strong multisensory responses (Meredith and Stein, 1986; Stein and Meredith, 1993), we also modeled superior colliculus bimodal data and compared the results to the cortical bimodal neurons. Interestingly, data from bimodal neurons in superior colliculus do not follow a vector-like model but in aggregate behave much like the City Block (m = 1) model discussed above. Finally there has been much interest in Bayesian-like models of multisensory integration but some reluctance to apply these models to bimodal neurons. This reluctance turns out to be well founded – Bayesian Maximum Likelihood Estimation (MLE) modeling produces the wrong prediction for audiovisual variance and underestimates audiovisual firing rates.

RESULTS

Audiovisual summation in cortex resembles psychophysical data and is vector-like

One of our laboratories has gathered data on many bimodal multisensory neurons, with two-variable (audiovisual as a function of best-of-audio-or-visual responses) descriptions of each neural data set given in Meredith et al. (2012). Working with the full three-variable audio, visual and audiovisual database, we modeled, in the Minkowski framework, data for 50 facilitatory audiovisual bimodal neurons in superior colliculus and in 74 audiovisual bimodal neurons from two different cortical areas of the cat. For example, Figure 2 shows the data from Figure 1C with the missing weaker unisensory response data restored. Data in Figures 2, 3, and 4 are shown using a two-dimensional three-variable scheme – invented for the present study – which makes relationships between sensory input and combination data interpretable (in a way conducive to understanding three-variable experimental data from multiple neurons), while allowing easy visual comparisons between predicted and actual values (avoiding some of the difficulties encountered with three-dimensional perspective and contour plots). For example, in Figure 2, neural spike rates for audio, visual and audiovisual combinations are shown as a function of the audiovisual combinations. The audiovisual neural combination data is therefore constrained to fall on the Line-of-Unity (the diagonal multisensory data axis), which is a straight line in all coordinate systems. The separate audio and visual responses contributing to audiovisual responses are shown relative to the audiovisual responses. Minkowski models for the data are computed by plugging the actual spike rates for audio and visual responses into the Minkowski equation (Equation 2) and setting the exponent m to predict the audiovisual firing rate.

$$BimodalFiringRate = (VisualFiringRate^m + AudioFiringRate^m)^{1/m} \qquad \text{(Equation 2)}$$

The model predictions are fit to lines and those lines are plotted in the same space; the plot of models is also effectively a plot of predicted values as a function of actual data; any deviation from the actual neural data on the Line-of-Unity is conspicuous in these plots. The relationship between data and models and the relationship between various models are more obvious and more readily interpretable in this representation than in a 3D perspective plot or in a contour plot. The results of these calculations are shown in Figures 2 and 3, for audiovisual neurons in cat cortical areas PLLS and AES, respectively.

In Figures 2A and 2B, the best fit to the audiovisual spike rate data coincides with the diagonal for an m of 1.75. That is, if we plug m values and unisensory firing rates into Equation 2 and fit a line through the
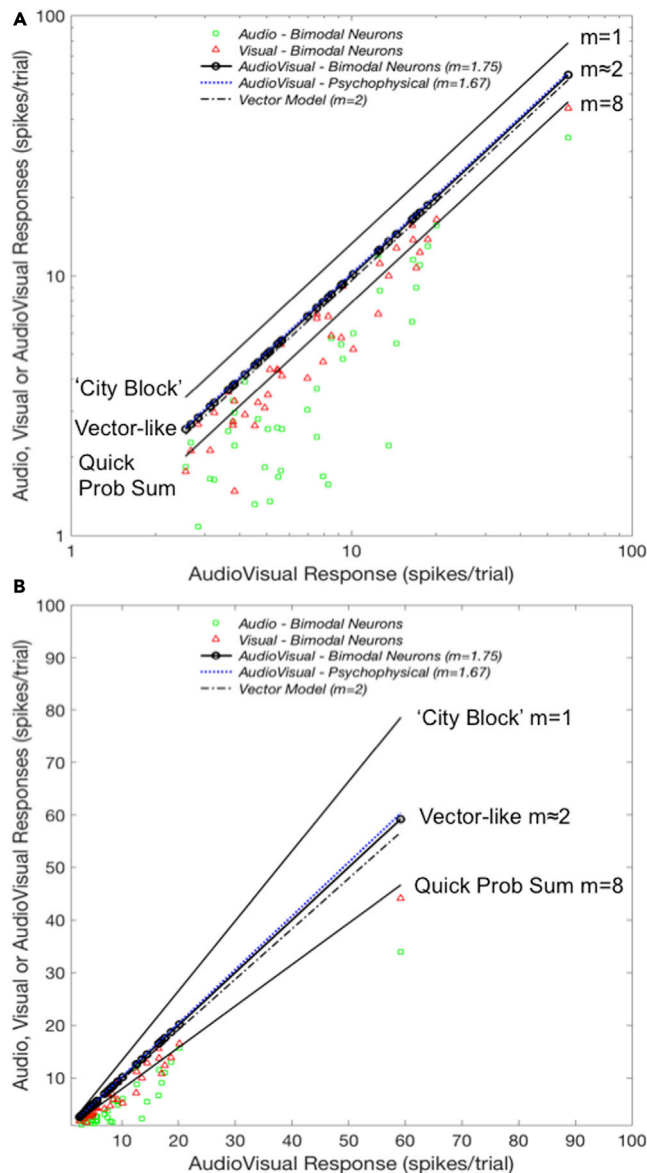
Figure 2. Audiovisual summation in PLLS neurons resembles perceptual data and is vector-like

Audiovisual response data for 40 cortical bimodal audiovisual neurons in cat PLLS (data from Figure 1C with omitted unisensory data restored). All neural spike counts in Figures 2, 3, 4, 5, and 6 are for 600 msec trials. Individual neural audio and visual responses are shown in relation to the combined audiovisual responses. Plots are mathematically designed to place audiovisual spike rates on the diagonal (neural data axis). The Equation 2 model for these data falls on the same diagonal if the exponent $m = 1.75$. Other theoretically interesting models (vector, city block, Quick's probability summation model and a model based on Schnupp et al. (2005) psychophysical data ($m = 1.67$) pop out as deviations from the diagonal. Each model is fit by taking the actual audio and visual responses of the PLLS cells and using Equation 2 with the desired model exponents to generate simulated neural responses.

(A) Logarithmic depiction spreads neural data out better for inspection.

(B) Linear depiction spreads out models to better illustrate differences between models. The neural data are not far from the well-known vector model ($m = 2$), but they are closer to Schnupp's psychophysically derived model ($m = 1.67$) than to any other theoretically based model.

predicted audiovisual values as a function of the actual audiovisual responses, we get a slope of 1.0 for an $m = 1.75$ for the 40 audiovisual PLLS neurons. (Area PLLS is a cat extrastriate visual cortex, located in Brodmann Area 19 and contains motion sensitive cells, somewhat similar to Area MT in macaque). Also plotted for comparison is the function for an m value of 1.67 found for Schnupp's psychophysical data, and several
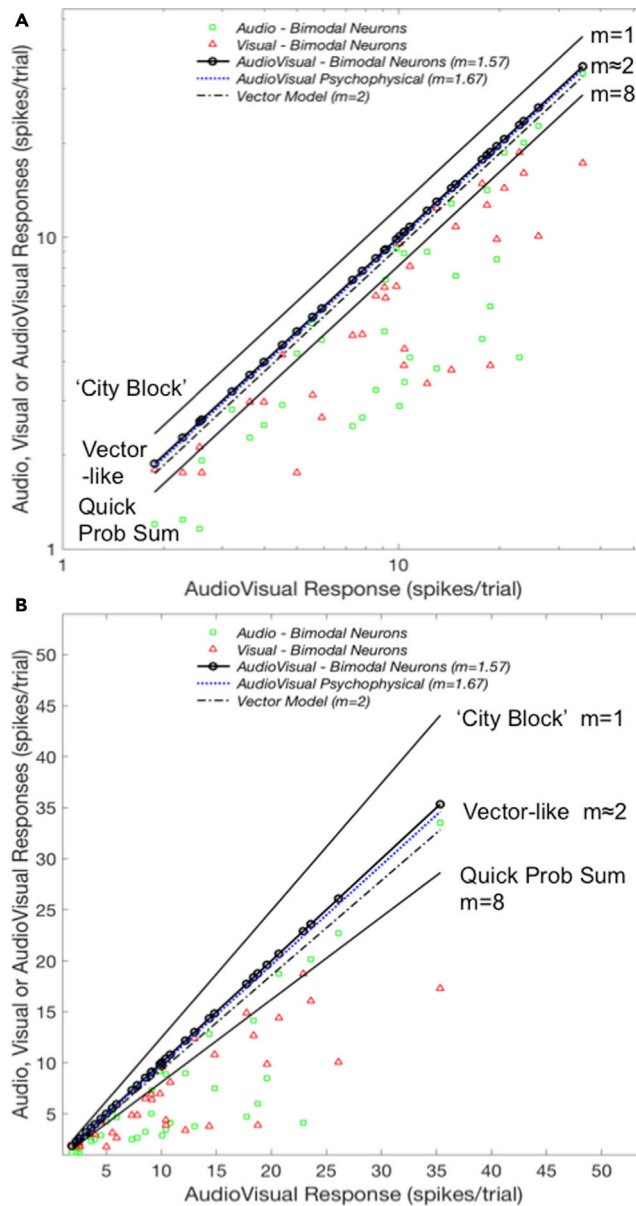
Figure 3. Audiovisual summation in AES neurons resembles perceptual data

Audiovisual firing rate data for 34 facilitatory cortical bimodal audiovisual neurons in cat AES, plotted as in Figure 2. The Equation 2 model for this data coincides with the neural audiovisual data diagonal if the exponent $m = 1.57$. Other theoretically interesting models pop out as deviations from the diagonal.

(A) Logarithmic depiction spreads neural data out better for inspection.

(B) Linear depiction spreads out models to better illustrate differences between models. The neural data are closer to Schnupp's psychophysically derived model ($m = 1.67$) than to any other theoretically based model.

theoretically interesting models: $m = 1$ (City Block metric), $m = 2$ (classic vector model) and $m = 8$ (Quick (1974) influential probability summation model), to situate the reader in the space for which Minkowski models generally occur. Similarly, in Figures 3A and 3B, we show the results for 34 audiovisual neurons in cat AES (multisensory association cortex), with the weaker unisensory response data restored. Here the Equation 2 model coincides with the neural audiovisual diagonal for an m value of 1.57. The psychophysical m of 1.67 ± 0.26 is similar to the estimates of m of 1.75 for PLLS neurons and 1.57 for AES neurons, suggesting that either or both neural populations could underlie psychophysical audiovisual detection thresholds. Indeed, if we assume that both PLLS and AES could be contributing to audiovisual detection,
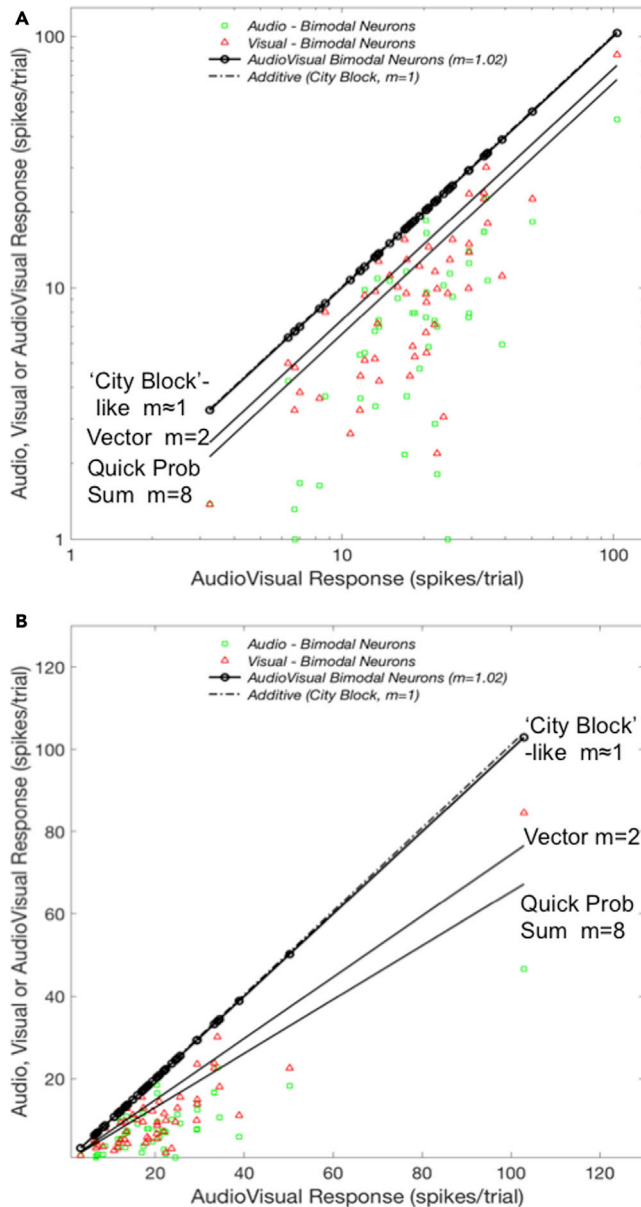
Figure 4. Audiovisual summation in cat superior colliculus resembles the City Block metric from experimental psychology

Audiovisual firing rate data for 50 facilitatory bimodal audiovisual neurons in cat superior colliculus, plotted as in Figures 2 and 3. The Equation 2 model for this data falls on the multisensory neural data axis if the exponent m = 1.02, an exponent that resembles the value of 1 used in the well-known 'City Block' modeling metric. Note that although the combined neural responses in Figures 2, 3, and 4 fall on the same diagonal neural data axis (as they must), that the Figures 2, 3, and 4 neural data axes are actually associated with different m values. Minkowski (m) exponents are computed from the relationship between unisensory and multisensory values, not from a plotting constraint.

(A) Logarithmic depiction spreads neural data out better for inspection; however the deviation of the City Block model from the neural data axis is barely visible.

(B) Linear depiction better spreads out models; the slight difference between the combined neural data axis and the City Block model can be seen here.

their weighted average m would be 1.67, remarkably consistent with the psychophysical result. However, if one were interested in a neural correlate of the more classic vector model (m = 2), then the PLLS cells, with an m estimate of 1.75 would be a closer match than the AES cells. An alternate way to characterize the m

value of the cell population would be to simply compute m for each individual neuron and to find the mean m value of the sample neurons. Given the unimodal and bimodal firing rates for each neuron, we can compute each neuron's m directly by root finding, e.g., by finding the zero-crossing of Equation 3.

$$AudioVisualFiringRate - (VisualFiringRate^m + AudioFiringRate^m)^{1/m} = 0 \qquad \text{(Equation 3)}$$

This alternate method yields estimates of $m = 2.07$ for the 40 PLLS neurons and $m = 1.86$ for the 34 AES cells, which are similar to the m of 2 for vector models. Indeed, if both neural populations are pooled the mean exponent m comes out as 1.97. However, since these are nonlinear systems and the average Minkowski exponent m value of a nonlinear ensemble may not be the most representative m characteristic of the population as a whole, we would lean more strongly on the slightly lower estimates for m produced by the fitting method employed in Figures 2 and 3.

Audiovisual summation in cat superior colliculus resembles the city block metric from experimental psychology

In psychology there are Minkowski models that use a 'City Block' (or 'Manhattan') metric (Attneave, 1950; Coombs et al., 1970; Garner, 1974; Landahl, 1945; Wuerger et al., 1995), in which $m = 1$ and the sensory components add linearly (as x and y distances would add on a street map if one were confined to traveling on north/south and east/west city streets). City Block metrics are often compared to Euclidean vector-like metrics ($m = 2$), where distances are subadditive because one can travel on a diagonal. In neural terms, a City Block metric would imply simple firing rate additivity (Landahl, 1945). In practice, all populations of bimodal cells seem to have both superadditive ($m < 1$) and subadditive ($m > 1$) neurons (Perrault et al., 2005; Meredith et al., 2012). As we have seen, in cortical populations subadditive neurons predominate and the audiovisual neural population as a whole approximate an m of 2. However, in superior colliculus, where bimodal cells were first discovered, many cells are superadditive, some strongly superadditive (Stein and Meredith, 1993). Estimates of the proportion of superadditive cells in superior colliculus range from about one-quarter to about one-half (Perrault et al., 2005; Meredith et al., 2012). In our sample, 66% of facilitatory audiovisual superior colliculus neurons were superadditive and their superadditivity ranged from slight ($m = 0.976$) to strong ($m = 0.287$). These superadditive neurons balance the also abundant subadditive cells so that the mean m of the 50 audiovisual superior colliculus bimodal neurons, computed from Equation 3, is actually quite close to additive ($m = 1.08 \pm 0.81$). Similarly, the Equation 2 model coincides with the audiovisual neural data diagonal for an m value of 1.02, surprisingly close to the exponent of 1 for additive City Block-like models. The approximation of cortical neurons to vector models and the approximation of superior colliculus neurons to City Block models are functionally and theoretically intriguing because that suggests that these neural regions process multisensory information in ways that are mathematically distinct (see Discussion). Further data are needed to evaluate this interesting dichotomy.

Bayesian inspired modeling: MLE predictions do not match audiovisual bimodal data

An influential school of thought in multisensory integration suggests that multisensory systems behave like ideal statistical estimators in the Bayesian sense: the multisensory responses can be modeled as a weighted average of the unisensory responses, with the weights determined by the inverse variances of the unisensory responses (Alias and Burr, 2004; Ernst and Banks, 2002). In principle, this will result in an unbiased estimate of the compound response and the variance of the multisensory response will be reduced relative to the unisensory responses. Much evidence favors this theory, some results do not support it and some theoretical work has gone into understanding the conditions that would result in non-ideal integration (Ernst, 2012). Because we have much data on neural variability and no data on neural Bayesian priors, the variant of Bayesian theory which is best suited to modeling spike count data is maximum likelihood estimation (MLE), which has widely been applied to sensory data (see Ernst, 2012 and Ma and Pouget, 2008 for reviews) and which is equivalent to a full Bayesian model for relatively flat priors. In the past, it has been suggested that bimodal neurons – especially superadditive bimodal neurons (which make up a substantial minority of our data) may be poor candidates for MLE (Ma and Pouget, 2008; Beck et al., 2008) but prior sensible results with MLE for visual-vestibular bimodal neurons (Morgan et al., 2008; Gu et al., 2008; Fetsch et al., 2010; Angelaki et al., 2012) suggest that it might be interesting to try. MLE makes two strong predictions about auditory-visual responses and their variability. First, audiovisual spike rates should be a weighted sum of audio and visual spike rates, with the weights determined by the variability of the audio and visual responses.

$$VA = k_v V + k_a A \qquad \text{(Equation 4)}$$

Where *VA* is the firing rate for bimodal stimulation, *V* is the firing rate for visual stimulation and *A* is the firing rate for auditory stimuli. If we define σ_V^2 as the variance in spike rate for visual stimulation and σ_A^2 as the variance in spike rate for auditory stimulation, the weights k_V and k_A are given as

$$k_v = (1/\sigma_V^2)/(1/\sigma_V^2 + 1/\sigma_A^2); \ k_a = (1/\sigma_A^2)/(1/\sigma_V^2 + 1/\sigma_A^2) \qquad \text{(Equation 5)}$$

Thus, MLE gives more weight to less variable responses. Note that this model yields weights that sum to 1 (a weighted average).

Second, MLE predicts that if the noise in the audio and visual channels is independent (an unlikely assumption for single neurons), then summing the audio and visual channels will result in result in lower audiovisual variance σ_{VA}^2.

$$\sigma_{VA}^2 = \sigma_V^2 \sigma_A^2/(\sigma_V^2 + \sigma_A^2) \qquad \text{(Equation 6)}$$

The variance prediction can immediately be disposed of. For each PLLS neuron, we have variance data based on 23 trials for each condition; for AES and SC neurons 25 and 16 trials, respectively, were available. MLE predicts that combined reliability should be greater than unimodal reliabilities, so variance (Equation 6) for response to audiovisual stimulation should be lower than either the variance for response to audio and visual stimulation. This is true for only 7% of tested neurons: 5 of 40 PLLS bimodal neurons, 2 of 34 AES bimodal neurons and 2 of 49 superior colliculus bimodal neurons (variance data for one superior colliculus neuron was missing).

The firing rate prediction is also not supported. Figure 5A shows the MLE predictions for audiovisual spike rates plotted against the observed PLLS audiovisual spike rates, on a cell-by-cell basis for the 40 neurons in our sample. Figure 5A also shows the Minkowski model for comparison. The MLE model (Equations 4 and 5) underestimates audiovisual bimodal spike rates. Based on a regression model, the MLE model firing rates are about 65 $\pm$ 2% (sd) of the observed neural values. Moreover, the Minkowski model outperforms the MLE model for 37/40 cells. For the 34 neurons from AES, the MLE prediction systematically underestimates the audiovisual combined response for each neuron (Figure 5B). Moreover, the Minkowski model outperforms the MLE model for 31/34 cells. The underestimate, obtained by linear regression, is about 63 $\pm$ 6% of the observed neural values. For the 49 neurons in superior colliculus, the MLE prediction underestimates the combined audiovisual response for each neuron (Figure 6). The underestimate, obtained by linear regression is about 42 $\pm$ 2% of the observed neural values. Moreover, the Minkowski model outperforms the MLE model for 43/49 cells. See Discussion for a treatment of what these results may imply.

DISCUSSION
Could bimodal neurons find potential uses in Bayesian-inspired neural models?
MLE models generate predictions that are poor fits to the actual neural data, but this may not be entirely disqualifying for some purposes. MLE models based on Equations 4, 5, and 6, are weighted averages that are limited to weights that sum to 1 and discount the more variable response (which is often the strongest response because variability is correlated with firing rate). Furthermore, in neural-based MLE models and in some MLE-inspired best-fit models, the weights (even if unconstrained) are frequently subadditive (Beck et al., 2008; Morgan et al., 2008; Gu et al., 2008; Fetsch et al., 2010; Angelaki et al., 2012). Because of this, it has been suggested that MLE models may not be well suited to enhanced response bimodal neurons, especially for strongly superadditive bimodal neurons (Ma and Pouget, 2008; Beck et al., 2008). This is a sensible argument, and the consistent underestimation of multisensory firing rate generated by the MLE models for our bimodal cells supports this. Certainly, the ability of the Minkowski model to represent both superadditive neurons and superadditive systems gives it versatility relative to Bayesian models. Similarly, the MLE model wrongly predicts that the variances of the multisensory response will be lower than the variances of the unisensory responses; in fact the multisensory response variance was almost always higher than at least one (114/123) and often both of the unisensory variances. Interestingly, given Ma and Pouget (2008) reservations, all of the superadditive cells violated the MLE variance prediction, but so did most of the subadditive cells. The MLE model assumes that noise in unisensory neural responses are independent; independence is unlikely to be obtained when computed on a neuron-by-neuron basis. The audio and visual inputs (and their noise) should be relatively independent, but when they are being processed by the same bimodal neuron, that neuron is using the same spike generating machinery and its associated noise
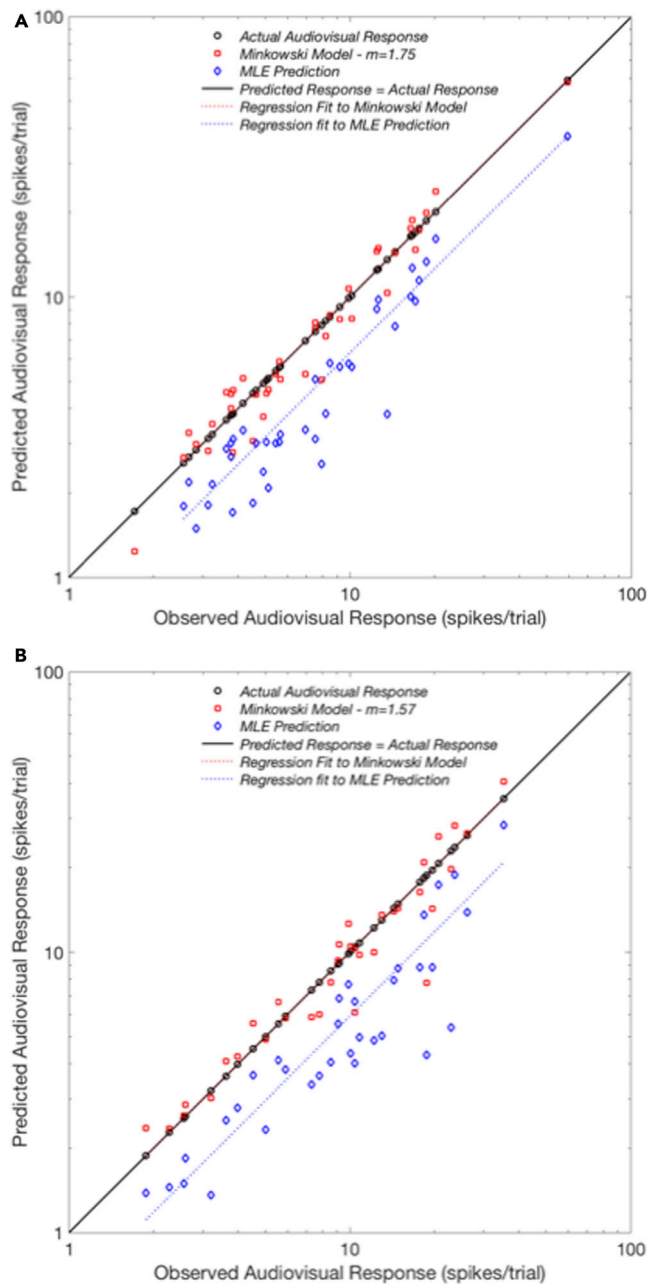
Figure 5. Comparison of Minkowski and Bayesian MLE models to actual cortical bimodal neuron data

Data are plotted as predicted audiovisual firing rates, as a function of actual audiovisual firing rates; the observed firing rate data (black) thus lies on a 45 deg. axis.

(A) PLLS neurons. The Minkowski model already depicted in Figure 2 (m = 1.75) is shown here with the simulated firing rates (red icons) it was derived from. A regression fit through this model coincides with the neural data axis (estimated response = actual response), with an r^2 value of 0.980. The maximum likelihood estimation (MLE) model, shown in blue, consistently (40/40 neurons) underestimates the actual firing rates. A regression fit through this MLE model suggests that on average the underestimate is about 65 $\pm$ 2% (sd) of the actual firing rates.

(B) AES neurons. The Minkowski model already depicted in Figure 3 (m = 1.57) is shown here with the simulated firing rates (red icons) produced by that model. A regression fit through this Minkowski simulation coincides with the neural data axis, with an r^2 value of 0.886. The MLE model, shown in blue, consistently (34/34) underestimates the actual data. A regression line fit through the simulated MLE data suggests that the underestimate is about 63 $\pm$ 6% of the actual firing rates.
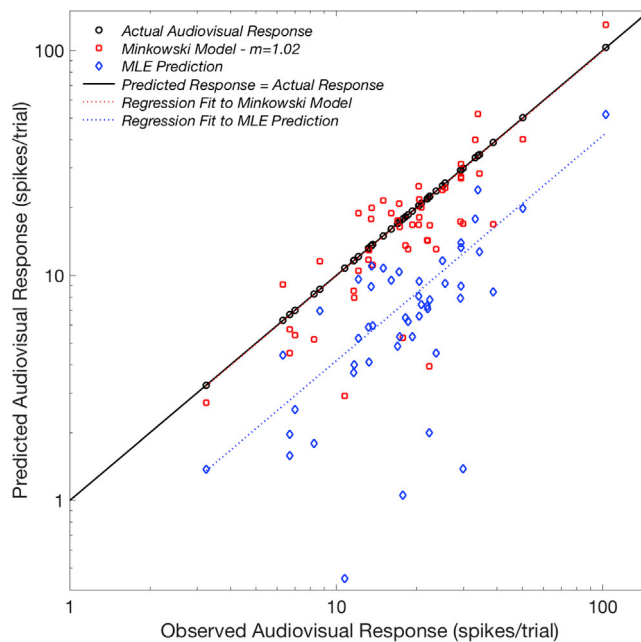
Figure 6. Comparison of Minkowski and Bayesian MLE models for 49 bimodal superior colliculus neurons
Data are plotted as predicted audiovisual firing rates, as a function of actual audiovisual firing rates; the observed firing rate data (black) thus lies on a 45 deg. axis. The Minkowski model already depicted in Figure 4 (m = 1.02) is shown here with the simulated firing rates (red icons) it was derived from. A regression model fit to the Minkowski simulation lies just underneath the neural data axis, where it can be seen as a red stippling, with an r^2 value of 0.815. The maximum likelihood estimation (MLE) prediction (shown in blue) consistently (49/49 neurons) underestimates the actual firing rate data; a regression line fit through the MLE simulated data suggests that the underestimate is about 42 $\pm$ 2% of the actual firing rates.

for both inputs. Indeed audio and visual spike rates computed on a neuron-by-neuron basis are highly correlated – sensitive neurons are sensitive to both inputs; insensitive neurons are insensitive to both inputs. Building a valid MLE model with audiovisual bimodal neurons might require pooling large numbers of neurons to obtain independence of pooled unisensory responses. However, this does not change the fact that the Figures 5 and 6 MLE firing rate underestimates are roughly proportional to observed neural responses and in neural theory proportionality is often sufficient. For example, in neural models, a simple sum of spike counts is sometimes used as a neural surrogate for the neurally more complicated averaging process (see De Valois et al., 1966 for an early example). Also, some Bayesian theorists do not believe that the brain is keeping track of variability in channels per se. Instead they have neural models of many variable interacting integrate-and-fire neurons. The Bayesian-like behavior of the network can be accounted for by attractor dynamics among variable network units. If this is the case, the proportionality found here might encourage trying realistic cortical bimodal neurons in Bayesian-inspired interacting neural network models like that of Beck et al. (2008). Alternately it may be enough in some non-rigorous Bayesian-inspired models to simply compensate for the underestimation.

Audiovisual summation benefits detection more than discrimination

There is another psychophysical study (To et al., 2010) that fits a Minkowski model to audiovisual discrimination data and got a somewhat greater value of m (2.56) than that obtained for Schnupp et al. (2005) audiovisual detection thresholds (i.e., audiovisual summation benefits detection more than discrimination). To et al. (2010) note that a vector model for discrimination would be expected if independent noisy channels combined for discrimination judgments; the combined signal-to-noise ratio would rise as the square root of the number of combined signals (see also Campbell and Green, 1965; Green and Swets, 1966). Since larger values of m imply less summation, To et al. considered factors that might result in less summation for cue combination in discrimination than a vector model would imply. They argue that if the channels were not quite independent, then their combined information would sum less efficiently than a vector sum,

resulting in a slightly higher *m*. They suggest that the inefficiency would be described by a modified vector sum that resembles the Law-of-Cosines.

$$AudioVisual = (Audio^2 + Visual^2 - 2*p*Audio*Visual)^{1/2} \qquad \text{(Equation 7)}$$

This variation of the Mahalanobis distance is used in perceptual similarity modeling (Ashby and Perrin, 1988). A related model was considered by Lehky (1983) for binocular and binaural summation. To et al. estimated a slight correlation *p* (between the audio and visual channels) of about 0.15 ± 0.02 would be sufficient to account for the deviation from a vector model for perceptual discrimination data. Neural correlations in this range have been reported by several studies (see for example Gawne and Richmond, 1993). This effect might also account for the discrepancy between the Minkowski exponents for discrimination and detection, if audiovisual discrimination used information in correlated cortical mechanisms less efficiently than did audiovisual detection.

Differences between cortical and superior colliculus bimodal neurons

When bimodal cells were first studied in superior colliculus, their most interesting characteristic was that some of these cells were very superadditive, with enormous response facilitations for multisensory combinations, compared with their strongest unisensory response (Meredith and Stein, 1986; Stein and Meredith, 1993). As discussed above, the average facilitatory bimodal neuron in PLLS or AES cortex is less facilitatory than the average superior colliculus neuron, and this is reflected by the percentage of superadditive cells ($m < 1$) found in each population. 66% of our audiovisual facilitatory superior colliculus bimodal neurons were superadditive by this criteria, compared to just 10% of PLLS bimodal neurons and 18% of AES bimodal neurons. A good review of differences between cortical and superior colliculus cells is Meredith et al. (2012). There is also a range of integrative behaviors seen within the superior colliculus (Perrault et al., 2005); it is likely that cortical cells vary similarly (Meredith et al., 2012). The extra enhancement found in superior colliculus neurons may be caused by cortical modulation of superior colliculus. Jiang et al. (2001) used cryogenic techniques to deactivate cortical feedback onto superior colliculus and found that much of the enhancement found in superior colliculus neurons goes away during cortical deactivation, while their separate unisensory responses are unaffected. Meredith et al. (2012) argue that there could be advantages to differences between facilitatory cortical and superior colliculus bimodal neurons. In superior colliculus, strong superadditivity could aid rapid sensory orienting responses to fleeting threats or opportunities. In cortex, where sensory representations could be distorted by strong superadditivity, weaker facilitation should be less problematic.

Multisensory interactions resemble color vision interactions

The similarity of the Minkowski models for audiovisual thresholds and bimodal spike rates indicate that the responses of cortical bimodal neurons are a likely neural correlate of multisensory perceptual sensitivity. Conversely, if we accept that cortical bimodal cells might underlie vector-like sensory information combinations, then the ubiquity of Minkowski models, and especially of vector models in sensory psychophysics, suggests that it might be useful to search for cortical neurons that could underlie Minkowski-like within-sense combinations, like those that have been neurally modeled for binocular or binaural summation (Lehky, 1983), but have never been rigorously compared to actual neurons that combine binaural or binocular signals. Another useful place to start might be in color vision interactions, especially interactions between the achromatic and hue signals. Color vision interactions have long been studied using nonlinear combination rules, including vector and other Minkowski models (Kruskal, 1964; Guth and Lodge, 1973; Ingling and Tsou, 1977; Guth et al., 1980; Billock, 1995; Eskew et al., 1999; Billock and Tsou, 2005; Zhou and Mel, 2008; Gheiratmand and Mullen, 2014; Gheiratmand et al., 2016).

Although it seems odd to start with cats integrating audio and visual signals and to end with the possibility that color vision interactions could work similarly, this is not as big a leap as it seems. Before audiovisual bimodal cells were most intensively studied for cat (e.g., Meredith and Stein, 1983; 1985; 1986; Wallace et al., 1992; Stein and Meredith, 1993), similar visual-infrared neurons had been found in rattlesnake optic tectum (Hartline et al., 1978; Newman and Hartline, 1981). In the literature, the rattlesnake's visual-infrared integration is considered a form of multisensory integration because the signals are transduced in two adjacent but separate sensory organs (eyes and facial pits) and are transmitted separately by two nerves (optic and trigeminal), before they are projected onto optic tectum bimodal neurons. But to think about visual-infrared integration functionally rather than anatomically, the act of transducing different parts of the

electromagnetic spectrum and combining that spectral information might reasonably be considered as a form of color vision.

Bimodal neurons are not the only relevant neurons to be considered in the contexts of color vision and multisensory interaction. Newman and Hartline (1981), in the same rattlesnake optic tectum layers where they found bimodal visual-infrared neurons, also found neurons that respond, say to visual, not to infrared, but fire harder when an infrared signal accompanies visual stimulation. In the literature, such neurons are now generally called subthreshold multisensory neurons because in contrast to the bimodal cells, the weaker input is modulatory but not by itself strong enough to evoke a spiking response. Recently the properties of these neurons have been studied extensively (Allman and Meredith, 2007; Allman et al., 2008, 2009; Billock and Tsou, 2014). Billock and Havig (2018) found that audiovisual subthreshold multisensory cells in cat visual cortex implement gated amplifications of visual responses. Billock and Havig also modeled several supra-threshold perceptual interactions that look like they are gated amplifications of the underlying perceptual strengths. Both the neural and perceptual interactions obey a simple power law with similar exponents (n ≈ 0.85)

$$AmplifiedResponse = a*UnamplifiedResponse^n \qquad \text{(Equation 8)}$$

Billock and Havig (2018) argue that these gated amplifier neurons could be the neural correlates of supra-threshold perceptual amplification and that their slightly compressive power law amplifications are consistent with the Principle of Inverse Effectiveness. The data modeled in the present paper complement this finding by providing an analogous (but vector-like) potential neural substrate for a threshold perceptual behavior (detection). It seems likely that multisensory perceptual sensitivity could be mediated by cortical bimodal neurons and that multisensory perceptual appearance could be mediated by subthreshold multisensory cortical neurons that do gated amplifications. The two most studied neural cell classes in cortical multisensory interactions (bimodal neuron and gated amplifier neurons) correspond to two important perceptual modes: threshold performance and suprathreshold appearance.

There are interesting analogs to this appearance/performance dichotomy in color vision. Guth and Lodge (1973) created a vector model of color vision in large part to account for detection thresholds for combinations of individually weak (subthreshold) color stimuli. Ingling and Tsou (1977) and Guth et al. (1980) extended this vector model to account for suprathreshold interactions in general, including the shape of the suprathreshold chromatic brightness function. This may have been unnecessary – Billock and Havig (2018) found that they could account for the shape of the chromatic brightness function as a hue-gated amplification of the achromatic luminance system (e.g., it followed Equation 8 with an r^2 of 0.993). Their model for chromatic brightness is very similar to their models for multisensory appearance and multisensory gated amplifier neurons (Billock and Havig, 2018), suggesting that there could exist hue-gated amplifier neurons that amplify suprathreshold achromatic responses into chromatic brightness percepts, much like the audio-gated-amplifier neurons in the cat visual system. But this begs the question about how to account for Guth and Lodge (1973) vector-like combinations of subthreshold color stimuli. An obvious possibility would be a class of bimodal-like neurons (or a simple neuronal network implementing Equation 1), which combines hue and luminance signals instead of audio and visual signals.

CONCLUSIONS

We created a coherent framework for comparing neural and perceptual information combination in multisensory integration. Mathematically, audiovisual integration in two populations of cortical bimodal neurons is strikingly similar to audiovisual integration in human perception, as assessed by psychophysical methods. Both electrophysiological and psychophysical data are in turn compatible with vector-like summation, a widely used model in cognitive science. The results suggest that cortical bimodal neurons likely underlie human audiovisual sensitivity and provide a neural mechanism for implementing vector-like summations – a neural mechanism that may generalize to other sensory systems like color vision. However, audiovisual bimodal neurons in cat superior colliculus, in aggregate, follow another theoretically interesting ('City Block') combination rule sometimes found in experimental psychology. And a previous study found a still different rule (a simple power law gated amplification; Equation 8) is followed for psychophysical appearance measures and by a different class of cortical neurons – subthreshold multisensory cells – with nearly identical parameterizations for electrophysiology and psychophysics (Billock and Havig, 2018). Taken together, the results suggest that the two most studied neural cell classes in cortical

multisensory interactions – gated amplifier neurons and facilitatory bimodal neurons – may provide neural substrates for two important perceptual modes: appearance-based and performance-based perception.

Limitations of the study

This study examined audiovisual data from 74 bimodal neurons in two areas of cat cortex and 50 bimodal neurons from cat superior colliculus. It would be useful to have additional datasets from the same neural areas to establish repeatability and generality. Also it would be useful to have similar data from nonhuman primates for comparison to human psychophysics, and to deliberately align the experimental designs so that equivalence can be drawn between human behavioral and animal neural data. The visual and auditory stimuli for the cat (Meredith et al., 2012) and human (Schnupp et al., 2005) experiments are comparable in spatiotemporal content and effect. Schnupp et al. (2005) used flashing spots of light. Meredith et al. (2012) used flashing spots of light (as well as moving spots and moving lines when they were effective stimuli for individual neurons); all of these stimuli provide abrupt spatial discontinuities and temporal modulation well suited to neurons found in both cat and primate cortex. Similarly, both Schnupp et al. and Meredith et al. used controlled bursts of broadband auditory noise as stimuli. Proposed primate experiments should be similarly well aligned with human psychophysical experiments. We would expect similar results from such primate experiments. The cat visual and auditory cortices are similar to macaque visual and auditory cortices with multiple functional subdivisions/hierarchies with similar cell types, similar binocular and binaural integration and similar cortical plasticity. Cat spatial vision is similar to macaque but is shifted toward lower spatial frequencies (De Valois and De Valois, 1991). Studies of cat auditory behaviors generally resemble human auditory behaviors (Populin and Lin, 1988), and cats are sensitive to all but the lowest frequencies that humans hear (Heffner and Heffner, 1985). Thus we would not expect great differences between cat and primate audio and visual responses to the stimuli used in these experiments. Finally, there do not seem to be significant reported differences between cat and human/primate multisensory combination. Until recently, the cat has been the chief experimental animal in sensory integration, especially at the neuronal level, and it has been shown that neurons and orienting behaviors follow the same general rules in behavioral multisensory interactions (including the principles of spatiotemporal coincidence and inverse effectiveness; Stein et al., 1988; Stein and Meredith, 1993). Cats (Stein et al., 1989) and humans (Lovelace et al., 2003; Noesselt et al., 2010; Rach et al., 2011) exhibit similar multisensory effects on detection and orienting behaviors, and these responses are influenced by cortical function (Wilkinson et al., 1996). Furthermore, humans demonstrate cross-modal suppression during attention tasks (Teder-Sälejärvi et al., 1999) and a category of multisensory neurons that exclusively exhibit cross-modal suppression have been identified in cat cortices (Dehner et al., 2004). Given that cats and humans/primates have similar sets of sensory receptors, demonstrate similar neural circuitry to evaluate these inputs, and exhibit similar basic behaviors based on those inputs, these similarities make it reasonable to compare human multisensory psychophysics and cat electrophysiology.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- RESOURCE AVAILABILITY
  - Lead contact
  - Materials availability
  - Data and code availability
- EXPERIMENTAL MODEL AND SUBJECT DETAILS
- METHODS DETAILS
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Neural data used in calculations
  - Minkowski modeling and maximum likelihood estimation modeling
- ADDITIONAL RESOURCES

AUTHOR CONTRIBUTIONS

JWHS provided psychophysical data from his laboratory and algorithmic support. MAM provided electrophysiological data from his laboratory. VAB and MJK analyzed and interpreted the data. VAB conceived the project and wrote the manuscript with inputs from the other authors.

DECLARATION OF INTERESTS

REFERENCES

Alias, D., and Burr, D. (2004). The ventriloquist effect results from near optimal crossmodal integration. Curr. Biol. 14, 257–262.

Allman, B.L., Keniston, L.P., and Meredith, M.A. (2008). Subthreshold auditory inputs to extrastriate visual neurons are responsive to parametric changes in stimulus quality: sensory-specific versus non-specific coding. Brain Res. 1242, 95–101.

Allman, B.L., Keniston, L.P., and Meredith, M.A. (2009). Not just for bimodal neurons anymore: the contribution of unimodal neurons to cortical multisensory processing. Brain Topogr. 21, 157–167.

Allman, B.L., and Meredith, M.A. (2007). Multisensory processing in 'unimodal neurons: cross-modal subthreshold auditory effects in cat extrastriate visual cortex. J. Neurophysiol. 98, 545–549.

Angelaki, D.E., Gu, Y., Fetsch, C.R., and Deangelis, G.C. (2012). In The New Handbook of Multisensory Integration, B.E. Stein, ed. (MIT Press), pp. 483–494.

Ashby, F.G., and Perrin, N.A. (1988). Towards a unified theory of similarity and recognition. Psychol. Rev. 95, 124–150.

Attneave, F. (1950). Dimensions of similarity. Am. J. Psych. 63, 516–556.

Beck, J.M., Ma, W.J., Kiani, R., Hanks, T., Churchland, A.K., Roitman, J., Shadlen, M.N., Latham, P.E., and Puget, A. (2008). Probabilistic population codes for Bayesian decision making. Neuron 60, 1142–1152.

Billock, V.A. (1995). The spectral sensitivity of the acuity criterion: effect of nonlinear summation of psychophysically isolated parvocellular receptive field centers. In Colour Vision Deficiencies XII, B.

Drum, ed. (Kluwer Academic Publishers), pp. 259–265.

Billock, V.A., and Havig, P.R. (2018). A simple power law governs many sensory amplifications and multisensory enhancements. Sci. Rep. 8, 1–8.

Billock, V.A., and Tsou, B.H. (2005). Sensory recoding via neural synchronization: integrating hue and luminance into chromatic brightness and saturation. J. Opt. .Soc. Am. A Opt. Image Sci. Vis. 22, 2289–2298.

Billock, V.A., and Tsou, B.H. (2014). Bridging the divide between sensory integration and binding theory: using a binding-like neural synchronization mechanism to model sensory enhancements during multisensory interactions. J. Cogn. Neurosci. 26, 1587–1599.

Campbell, F.W., and Green, D.G. (1965). Monocular versus binocular visual acuity. Nature 208, 191–192.

Coombs, C.H., Dawes, R.M., and Tversky, A. (1970). Mathematical Psychology (Prentice Hall).

Dehner, L.R., Keniston, L.P., Clemo, H.R., and Meredith, M.A. (2004). Cross-modal circuitry between auditory and somatosensory areas of the cat anterior ectosylvian sulcal cortex: a 'new' inhibitory form of multisensory convergence. Cereb. Cortex 14, 387–403.

De Valois, R.L., Abramov, I., and Jacobs, G.H. (1966). Analysis of response patterns of LGN cells. J. Opt. Soc. Am. A 56, 966–977.

De Valois, R.L., and De Valois, K.K. (1991). Spatial Vision (Oxford University Press).

Ernst, M. (2012). Optimal multisensory integration: assumptions and limits. In The New Handbook of Multisensory Processes, B.E. Stein, ed. (MIT Press), pp. 527–543.

Ernst, M., and Banks, M.S. (2002). Humans integrate visual and haptic information in a statistcally optimal fashion. Nature 415, 429–443.

Eskew, R.T., McLellan, J.S., Giulianini, F., Gegenfurtner, K., and Sharpe, L.T. (1999). Chromatic detection and discrimination. In Color Vision: From Genes to Perception, K.R. Gegenfurtner and L.T. Sharpe, eds. (Cambridge University Press), pp. 345–368.

Fetsch, C.R., DeAngelis, G.C., and Angelaki, D.E. (2010). Visual-vestibular cue integration for heading perception: applications of optimal cue integration theory. Eur. J. Neurosci. 31, 1721–1729.

Forsythe, G.E., Malcolm, M.A., and Moler, C.B. (1977). Computer Methods for Mathematical Computations (Prentice-Hall).

Garner, W.R. (1974). The Processing of Information and Structure (Lawrence Erlbaum).

Gawne, T.J., and Martin, J.M. (2002). Responses of primate cortical V4 neurons to simultaneously presented stimuli. J. Neurophysiol. 88, 1128–1135.

Gawne, T.J., and Richmond, B.J. (1993). How independent are the messages carried by adjacent inferior temporal cortical neurons? J. Neurosci. 13, 2758–2771.

Gheiratmand, M., Cherniawsky, A.S., and Mullen, K.T. (2016). Orientation tuning of binocular summation: a comparison of colour to achromatic contrast. Sci. Rep. 6, 25692.

Gheiratmand, M., and Mullen, K.T. (2014). Orientation tuning in human colour vision at detection threshold. Sci. Rep. 4, 4285.

Green, D.M., and Swets, J.A. (1966). Signal Detection Theory and Psychophysics (Wiley).

Gu, Y., Angelaki, D.E., and DeAngelis, G.C. (2008). Neural correlates of multi-sensory cue integration in macaque area MSTd. Nat. Neurosci. *11*, 1201–1210.

Guth, S.L., and Lodge, H.R. (1973). Heterochromatic additivity, foveal spectral sensitivity, and a new color model. J. Opt. Soc. Am. *63*, 450–462.

Guth, S.L., Massof, R.W., and Benzschawel, T. (1980). Vector model for normal and dichromatic color vision. J. Opt. Soc. Am. *70*, 197–212.

Hartline, P.H., Kass, L., and Loop, M.S. (1978). Merging of modalities in the optic tectum: infrared and visual integration in rattlesnakes. Science *199*, 1225–1229.

Heffner, R.S., and Heffner, H.E. (1985). Hearing range of the domestic cat. Hearing Res. *19*, 85–88.

Horn, H.F., and Hill, R.M. (1966). Responsiveness to sensory stimulation of units in the superior colliculus and subadjacent tectotegmental regions of the rabbit. Exp. Neurol. *14*, 199–223.

Ingling, C.R., Jr., and Tsou, B.H.P. (1977). Orthogonal combination of the three visual channels. Vis. Res. *17*, 1075–1082.

Jiang, W., Wallace, M.T., Jiang, H., Vaughan, J.W., and Stein, B.E. (2001). Two cortical areas mediate multisensory integration in superior colliculus neurons. J. Neurophysiol. *85*, 506–522.

King, A.J., and Palmer, A.R. (1985). Integration of visual and auditory information in bimodal neurons in the Guinea-pig superior colliculus. Exp. Brain Res. *60*, 492–500.

Kruskal, J.B. (1964). Multidimensional scaling by optimizing goodness of fit to a nonmetric hypothesis. Psychometrika *29*, 1–27.

Landahl, H.D. (1945). Neural mechanisms for the concepts of difference and similarity. Bull. Math. Biophys. *7*, 83–88.

Lehky, S.R. (1983). A model of binocular brightness and binaural loudness perception in humans with general applications to nonlinear summation of sensory inputs. Biol. Cybern. *49*, 89–97.

Lovelace, C.T., Stein, B.E., and Wallace, M.T. (2003). An irrelevant light enhances auditory detection in humans: a psychophysical analysis of multisensory integration in stimulus detection. Cogn. Brain Res. *17*, 447–453.

Ma, W.J., and Pouget, A. (2008). Linking neurons to behavior in multisensory perception: a computational review. Brain Res. *1242*, 4–12.

Mavrides, C.M. (1970). Selective attention and individual preferences in judgmental responses to multifeature patterns. Psychon. Sci *21*, 67–68.

Meredith, M.A., Allman, B.L., Keniston, L.P., and Clemo, H.R. (2012). Are bimodal neurons the same throughout the brain? In The Neural Bases of Multisensory Processes, M.M. Murray and M.T. Wallace, eds. (CRC Press), pp. 51–64.

Meredith, M.A., and Stein, B.E. (1983). Interactions among converging sensory inputs in the superior colliculus. Science *221*, 389–391.

Meredith, M.A., and Stein, B.E. (1985). Descending efferents of the superior colliculus relay integrated multisensory information. Science *227*, 657–659.

Meredith, M.A., and Stein, B.E. (1986). Visual, auditory, and somatosensory convergence in cells in the superior colliculus result in multisensory integration. J. Neurophysiol. *56*, 640–662.

Minkowski, H. (1891). Uber die positiven quadratischen Formen und über kettenbruchähnliche Algorithmen. [On positive quadratic forms and on algorithms suggesting continued fractions. J. für die Reine Angew. Mathematik (Crelles Journal) *1891*, 278–297, Reprinted in Minkowski, H. (1967) Gesamelte Abhandlungen [Collected Treatises], Vol. I. (Chelsea.).

Minkowski, H. (1910). Geometre Der Zahlen [The Geometry of Numbers] (Druck und Verlag von Teubner)).

Morgan, M.L., DeAngelis, G.C., and Angelaki, D.E. (2008). Multisensory integration in macaque visual cortex depends on cue reliability. Neuron *59*, 662–673.

Mysore, S.P., Asadollahi, A., and Knudsen, E. (2011). Signaling of the strongest stimulus in owl optic tectum. J. Neurosci. *31*, 5186–5196.

Newman, E.A., and Hartline, P.H. (1981). Integration of visual and infrared information in bimodal neurons in the rattlesnake optic tectum. Science *213*, 789–791.

Noesselt, T., Tyll, S., Boehler, C.N., Budinger, E., Heinze, H.J., and Driver, J. (2010). Sound-induced enhancement of low-intensity vision: multisensory influences on human sensory-specific cortices and thalamic bodies relate to perceptual enhancement of visual detection sensitivity. J. Neurosci. *30*, 13609–13623.

Oleksiak, A., Klink, P.C., Postma, A., van der Ham, I.J., Lankheet, M.J., and van Wezel, R.J. (2011). Spatial summation in macaque parietal area 7a follows a winner-take-all rule. J. Neurophysiol. *105*, 1150–1158.

Perrault, T.J., Vaughan, J.W., Stein, B.E., and Wallace, M.T. (2005). Superior colliculus neurons use distinct operational modes in the integration of multisensory stimuli. J. Neurophysiol. *93*, 2575–2586.

Quick, R.F. (1974). A vector-magnitude model of contrast detection. Kybernetik *16*, 65–67.

Rach, S., Diederich, A., and Colonius, H. (2011). On quantifying multisensory interaction effects in reaction time and detection rate. Psych. Res. *75*, 77–94.

Schnupp, J.W., Dawe, K.L., and Pollack, G.L. (2005). The detection of multisensory stimuli in an orthogonal sensory space. Exp. Brain Res. *162*, 181–190.

Shepard, R.N. (1964). Attention and the metric structure of the stimulus space. J. Math. Psych. *1*, 54–87.

Shepard, R.N. (1987). Toward a universal law of generalization for psychological science. Science *237*, 1317–1323.

Stein, B.E., Honeycutt, W.S., and Meredith, M.A. (1988). Neurons and behavior: the same rules of multisensory integration apply. Brain Res. *448*, 355–358.

Stein, B.E., and Meredith, M.A. (1993). The Merging of the Senses (MIT Press).

Stein, B.E., Meredith, M.A., Huneycutt, W.S., and McDade, L. (1989). Behavioral indices of multisensory integration: orientation to visual cues is affected by auditory stimuli. J. Cogn. Neurosci. *1*, 12–24.

Teder-Sälejärvi, W.A., Münte, T.F., Sperlich, F.J., and Hillyard, S.A. (1999). Intra-modal and cross-modal spatial attention to auditory and visual stimuli. An event-related brain potential study. Cogn. Brain Res. *8*, 327–343.

To, M.P.S., Baddeley, R.J., Troscianko, T., and Tolhurst, D.J. (2010). A general rule for sensory cue summation: evidence from photographic, musical, phonetic and cross-modal stimuli. Proc. R. Soc. B Biol. Sci. *278*, 1365–1372.

Wallace, M.T., Meredith, M.A., and Stein, B.E. (1992). Integration of multiple sensory modalities in cat cortex. Exp. Brain Res. *91*, 484–488.

Wilkinson, L.K., Meredith, M.A., and Stein, B.E. (1996). The role of anterior ectosylvian cortex in cross-modality orientation and approach behavior. Exp. Brain Res. *112*, 1–10.

Wuerger, S.M., Maloney, L.T., and Krauskopf, J. (1995). Proximity judgments in color space: tests of a Euclidean color geometry. Vis. Res. *35*, 827–835.

Zahar, Y., Reches, A., and Gutfreund, Y. (2009). Multisensory enhancement in the optic tectum of the barn owl: spike count and spike timing. J. Neurophysiol. *101*, 2380–2394.

Zhou, C.H., and Mel, B.W. (2008). Cue combination and color edge detection in natural scenes. J. Vis. *8*, 1–25.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
No deposited data. See "Other" for provided data		
Software and algorithms		
MATLAB Version 2018b	Mathworks	2018b
	Natick, MA USA	
Other		
Reanalyzed data from Meredith et al. (2012) by courtesy	M.A. Meredith	
of the first author.	Virginia Commonwealth U.	
Reanalyzed data from Schnupp et al. (2005) by courtesy	J.W. Schnupp	
of the first author.	City University of	
	Hong Kong	

RESOURCE AVAILABILITY

Lead contact

Vincent A. Billock; vincent.billock.ctr@us.af.mil

Materials availability

Not applicable (computational study)

Data and code availability

MATLAB code used to analyze the data, and the neural data embedded in the code, can be obtained from VAB.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

(not applicable – computational study)

METHODS DETAILS

This paper introduces a new method for plotting multisensory data. It is standard practice in the multisensory literature to plot bimodal firing rates as a function of the strongest unimodal response. This practice was adopted to illustrate the enhanced responses of bimodal cells to bimodal inputs (multisensory facilitation) but is not helpful for understanding how inputs combine. For our purposes, we created the plotting method used in Figures 2, 3, and 4 to emphasize how information combines. Alternatives like contour plots were rejected as being difficult to interpret and three-dimensional prospective plots were rejected because it is difficult to show differences between models in that format. Because audio, visual and audiovisual responses are plotted as a function of audiovisual responses, the audiovisual responses plot on the diagonal (even if logarithmically transformed) and the separate audio and visual responses plot below the diagonal that they combine to yield. An alternative method would have been to plot audio, visual and audiovisual responses as functions of a fourth variable. In this case, no convenient fourth variable presented itself and the alternative of creating a dummy variable (like say cell number) was rejected as unnecessarily convoluted and less elegant than the method we employed. To plot Minkowski models in this space we used the actual measured firing rates of each cell for separate audio and visual stimulation in Equation 2 and simulated the combined audiovisual firing rates that would arise for various values of the Minkowski exponent m. Lines were then fit to the resulting simulated (Equation 2)

audiovisual firing rates for various theoretically and experimentally interesting values of the Minkowski exponent (m = 1, 1.67, 2, 8) and for the best fit Minkowski model to the bimodal neuron data. See Quantification and Statistical Analysis, below for more details.

QUANTIFICATION AND STATISTICAL ANALYSIS

Neural data used in calculations

All three variable (audio, visual and audiovisual) neural response data were taken from our earlier study (Meredith et al., 2012), which had considered and published a reduced two-variable description and analysis. In brief, the relevant methodological details of Meredith et al. (2012) are as follows: Cats were surgically fitted with recording wells and were anesthetized during this procedure; recording sessions took place 7-10 days later. Animals were anesthetized for recordings with a mixture of ketamine and acepromazine. Spontaneous movements were controlled with a muscle relaxant. Animals were intubated by mouth and maintained on a ventilator. Animals were fixed in place by attachment to the recording well implant. Each cat neuron was tested for effectiveness of stimulation with a variety of auditory stimuli and an effective response-eliciting stimulus for that neuron (either white noise or a digital waveform) was employed in data collection. Visual stimuli were moving bars of light projected on a translucent hemisphere in front of the animal. A combination of movement velocity/direction and bar size/orientation/luminosity that produced an effective response for that neuron was used for data collection. Data was collected using a glass insulated tungsten electrode, which was advanced with a hydraulic micro-drive. Depth of electrode was recorded for each studied neuron. Recording tracts were reconstructed histologically from fixed tissue sections taken from euthanized animals. Large numbers of audiovisual multisensory cells were found in three areas of cat brain: superior colliculus, cortical area PLLS (posterolateral suprasylvian cortex), and cortical area AES (anterior ectsylvian sulcus). Each stimulus was repeated 16 times for each superior colliculus neuron, 23 times for each PLLS neuron and 25 times for each AES neuron. An inter-stimulus interval of 7-15 seconds was used to avoid habituation. There were 49 audiovisual neurons found in cat cortical area PLLS, a visual area in Brodmann Area 19 that may correspond roughly to area MT in primates. Of these, eight were suppressive and one behaved like a MAX cell (Mavrides, 1970; Gawne and Martin, 2002; Mysore et al., 2011; Oleksiak et al., 2011; MAX neurons are theoretically interesting, and will be modeled in a separate study). The remaining 40 enhancing bimodal cells were analyzed here. There were 40 audiovisual neurons found in cat cortical area AES, a multisensory area found in association cortex. Of these, four were suppressive neurons and two behaved like MAX operators. The remaining 34 enhancing bimodal cells were analyzed here. Finally, there were 58 audiovisual neurons found in cat superior colliculus. Of these, two were suppressive neurons and six behaved more like subthreshold multisensory neurons (see Allman and Meredith (2007), Allman et al. (2008), Billock and Tsou (2014), and Billock and Havig (2018) for background on these multisensory but not bimodal neurons). The remaining 50 cells' bimodal responses were analyzed here.

Minkowski modeling and maximum likelihood estimation modeling

To find the Minkowski exponent m that best corresponds to a population of bimodal neurons' audiovisual responses, we used the actual measured firing rates of each cell for separate audio and visual stimulation in Equation 2 and simulated the combined audiovisual firing rates that would arise for various values of the Minkowski exponent m. Lines were then fit to the resulting simulated (Equation 2) audiovisual firing rates (as a function of actual audiovisual firing rate) for various values of the Minkowski exponent until we obtained a fitted line with the same slope (1.0) as the diagonal. This model was adopted as the characteristic Minkowski model for the group of neurons under study. Other plotted models were fitted similarly. For example, for a vector model, an m of 2 would be used in Equation 2 in conjunction with the actual audio and visual firing rates to simulate the audiovisual combined firing rate for a vector–like system. The simulated combined firing rates (as a function of the actual neural firing rates) are fit to a straight line and only the straight lines are plotted in Figures 2, 3, and 4 (to avoid cluttering the graphs with many simulated neural firing rates). However, the best fit Minkowski and Maximum Likelihood Estimation models – with their simulated audiovisual values – are shown in Figures 5 and 6. We also computed Minkowski exponents for individual neurons, to quantify the superadditivity and subadditivity of cells in each area of the brain. For each neuron a Minkowski exponent was computed by finding the root of Equation 3. A hybrid root finding algorithm was used that combined bisection, secant and inverse quadratic interpolating methods (Forsythe et al., 1977). Maximum Likelihood Estimation calculations (Equations 4, 5, and 6) were computed using mean firing

rates and associated variances described above with one exception: one Superior Colliculus neuron was excluded from the calculation shown in Figure 6 because missing auditory variance data precluded the MLE calculation for this neuron.

ADDITIONAL RESOURCES

(None – Computational Study)