

Bayesian Inference Explains Perception of Unity and Ventriloquism Aftereffect: Identification of Common Sources of Audiovisual Stimuli

Yoshiyuki Sato

yoshi@sat.t.u-tokyo.ac.jp

Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, University of Tokyo, Tokyo 277-8562, and Aihara Complexity Modeling Project, ERATO, Japan Science and Technology Agency, Tokyo 151-0064, Japan

Taro Toyoizumi

taro.toyoizumi@brain.riken.jp

RIKEN Brain Science Institute, Saitama 351-0198, Japan, and Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, University of Tokyo, Tokyo 277-8562, Japan

Kazuyuki Aihara

aihara@sat.t.u-tokyo.ac.jp

Institute of Industrial Science, University of Tokyo, Tokyo 153-8505, Japan; Aihara Complexity Modeling Project, ERATO, Japan Science and Technology Agency, Tokyo 151-0064, Japan, and Department of Complexity Science and Engineering, Graduate School of Frontier Sciences, University of Tokyo, Tokyo 277-8562, Japan

We study a computational model of audiovisual integration by setting a Bayesian observer that localizes visual and auditory stimuli without presuming the binding of audiovisual information. The observer adopts the maximum a posteriori approach to estimate the physically delivered position or timing of presented stimuli, simultaneously judging whether they are from the same source or not. Several experimental results on the perception of spatial unity and the ventriloquism effect can be explained comprehensively if the subjects in the experiments are regarded as Bayesian observers who try to accurately locate the stimulus. Moreover, by adaptively changing the inner representation of the Bayesian observer in terms of experience, we show that our model reproduces perceived spatial frame shifts due to the audiovisual adaptation known as the ventriloquism aftereffect.

1 Introduction ---

Audiovisual integration plays an important role in our perception of external events. When light and sound are emitted from an external event, we

take both of them into account to obtain as much information as possible, for example, the location and the timing of the event.

The well-known ventriloquism effect is a good example of audiovisual integration: an auditory stimulus is perceived to be biased toward the corresponding visual stimulus when presented in temporal coincidence (Howard & Templeton, 1996). Even simple audiovisual stimuli such as a small spot of light and a beep are known to cause a ventriloquism effect (Alais & Burr, 2004; Bertelson & Radeau, 1981; Hairston et al., 2003; Slutsky & Recanzone, 2001; Wallace et al., 2004).

In general, when conflicting multisensory stimuli are presented through two or more modalities, the modality having the best acuity influences the other modalities (the modality appropriateness hypothesis) (Welch & Warren, 1980). According to this hypothesis, a visual stimulus influences the localization of an auditory stimulus because vision has a greater resolution than hearing. In addition, it has been shown that a tactile location is also modified by a visual stimulus (Pavani, Spence, & Driver, 2000). In the case of temporal information, audition is more accurate than vision; the perceived arrival time of a visual event can be attracted to the sound arrival time (temporal ventriloquism) (Bertelson & Aschersleben, 2003; Morein-Zamir, Soto-Faraco, & Kingstone, 2003).

The ventriloquism effect can be understood as a near-optimal integration of conflicting auditory and visual stimuli. Alais and Burr (2004) computed a maximum a posteriori (MAP) estimation (or an equivalent maximum likelihood estimation in their model) of the location of the assumed common source of audiovisual stimuli. They found that their MAP estimator predicted the experimental ventriloquism effect very well.

However, binding audiovisual information is a good strategy only when the light and the sound have a common source. Binding visual and auditory stimuli even when they do not have a common source may impair our accurate perception of external events. Therefore, the judgment of the common sources of audiovisual stimuli plays a crucial role in audiovisual integration. This idea has been termed the unity assumption (Welch & Warren, 1980) or pairing (Epstein, 1975). The judgment on whether the presented audiovisual stimuli share the same source depends on their spatial and temporal discrepancies (Lewald & Guski, 2003). This result is reasonable because the visual and auditory stimuli that arrive close in space and time are likely to originate from a common event. Indeed, the strength of the ventriloquism effect decreases as the spatial and temporal disparity increases (Wallace et al., 2004). Moreover, the ventriloquism effect can be increased by cognitively compelling a common cause of perception of the auditory and visual events (Radeau, 1994). Putting these results together, the ventriloquism effect is strongly related to the perception of whether audiovisual signals are from the same source.

Several studies have modeled audiovisual interactions using the Bayesian inference (Battaglia, Jacobs, & Aslin, 2003; Ernst & Banks, 2002;

Witten & Knudsen, 2005). In these models, however, it is assumed that auditory and visual stimuli are integrated from the beginning. In this letter, we set a Bayesian observer that uses Bayesian inference to localize audiovisual stimuli without presuming the binding of the stimuli. Hence, the observer estimates the position or the onset time of the stimuli, judging whether the auditory and visual signals are from the same source.

In particular, we try to reproduce the following experimental results related to the ventriloquism effect and the perception of unity. There is a strong correlation between the ventriloquism effect and the spatial unity perception, that is, the perception of whether audiovisual stimuli are from the same place (Wallace et al., 2004). Another finding is that the perception of spatial unity depends not only on the spatial discrepancy but also on the temporal discrepancy between visual and auditory stimuli (Slutsky & Recanzone, 2001). Conversely, the perception of simultaneity is modulated by the spatial discrepancy (Bertelson & Aschersleben, 2003; Lewald & Guski, 2003; Zampini, Guest, & Shore, 2005).

Another result that we wish to reproduce in this letter is the effect of adaptation on the perception of audiovisual stimuli. Audiovisual localization is known to exhibit an adaptation phenomenon. After a prolonged exposure to simultaneous audiovisual stimuli with spatial disparity, the disparity that strongly produces the spatial unity perception shifts toward the presented disparity (Recanzone, 1998). The aftereffect can be observed not only in the audiovisual localization but also in the unimodal auditory localization. This effect is known as the ventriloquism aftereffect (Canon, 1970; Lewald, 2002; Radeau & Bertelson, 1974; Recanzone, 1998). The localization of an auditory stimulus shifts in the direction of a relative visual stimulus position during the adapting period. Visual localization is also modified by adapting stimuli; however, the magnitude of this effect is smaller than that of the auditory localization (Radeau & Bertelson, 1974). This asymmetry between vision and audition is also believed to be due to the difference in spatial resolutions. We model this effect of adaptation by changing the model conditional distributions during each presentation of a pair of auditory and visual stimuli.

Combining the above factors, we show that our model reproduces mainly the following three experimental results:

- A strong relationship between the ventriloquism effect and spatial unity perception
- A temporal influence on spatial perception and spatial influence on temporal perception
- Spatial audiovisual adaptation

In section 2, we introduce adaptive Bayesian inference and apply it to the estimation of audiovisual stimuli. In section 3, on the basis of numerical simulations, we show how our model reproduces experimental results on the ventriloquism effect, the perception of spatial unity, and the

ventriloquism aftereffect. Section 4 summarizes the results and discusses the theoretical and psychological plausibility of our model.

2 Modeling Audiovisual Interaction by Bayesian Inference _____

Bayesian models are shown to be one of the most powerful methods for understanding human psychophysical experiments on multisensory integration (Deneve & Pouget, 2004; Knill & Pouget, 2004; Körding & Wolpert, 2004). This method provides the optimal means to combine different kinds of information from different sensory pathways with different resolutions. Biologically plausible implementations of the method and their relations to the activity of neural circuits have also been investigated (Deneve, Latham, & Pouget, 2001; Rao, 2004). In the following section, we introduce a Bayesian model that estimates the location or the arrival time of audiovisual stimuli.

2.1 Bayesian Estimation Paradigm. For simplicity, we begin with a spatial task that does not have any temporal factors. Later, we also take temporal factors into consideration. The locations encoded in the early sensory systems are not exactly the same as the physically delivered ones; this is due to the sensory noise in the information processing pathways. Given the noisy sensory input, we assume that the Bayesian observer estimates a true location by using Bayesian inference (Körding & Wolpert, 2004; Witten & Knudsen, 2005). Further, we include the inference of whether audiovisual stimuli originate from the same event.

Given the sensory noises in auditory and visual information pathways, the perceived locations of the light and sound, x_V and x_A , respectively, are shifted from their original positions, x_{Vp} (light) and x_{Ap} (sound). Variable ξ is a binary variable that indicates whether visual and auditory stimuli originate from the same source ($\xi = 1$) or not ($\xi = 0$). The Bayesian observer uses the maximum a posteriori (MAP) rule for inference, that is, the observer determines its estimate such that the posterior probability is maximized, given the perceived locations. The posterior probability is calculated by integrating $P(x_{Vp}, x_{Ap}, \xi | x_V, x_A)$ with respect to unestimated parameters; for example, when only the location of an auditory stimulus is to be estimated, its MAP estimator \hat{x}_{Ap} is the one that maximizes the posterior probability $P(x_{Ap} | x_V, x_A)$, where the “hat” represents an estimator.

From Bayes’ formula, we obtain

$$P(x_{Vp}, x_{Ap}, \xi | x_V, x_A) = \frac{P(x_V, x_A | x_{Vp}, x_{Ap}, \xi) P(x_{Vp}, x_{Ap}, \xi)}{P(x_V, x_A)}, \quad (2.1)$$

where $P(x_V, x_A | x_{Vp}, x_{Ap}, \xi)$ is the conditional probability that the perceived positions are x_V and x_A given the physical parameters x_{Vp} , x_{Ap} , and ξ . We assume that the sensory noises in visual and auditory pathways are independent and are also independent of ξ . Hence, the first term on the

right-hand side in equation 2.1 is given by

$$\begin{aligned} P(x_V, x_A|x_{Vp}, x_{Ap}, \xi) &= P(x_V, x_A|x_{Vp}, x_{Ap}) \\ &= P(x_V|x_{Vp})P(x_A|x_{Ap}). \end{aligned} \tag{2.2}$$

The visual and auditory noises $P(x_V|x_{Vp})$ and $P(x_A|x_{Ap})$ are initially set according to gaussian distributions with mean zero, that is,

$$P(x_V|x_{Vp}) = \frac{1}{\sqrt{2\pi}\sigma_{sV}} \exp\left(-\frac{(x_V - x_{Vp})^2}{2\sigma_{sV}^2}\right), \tag{2.3}$$

$$P(x_A|x_{Ap}) = \frac{1}{\sqrt{2\pi}\sigma_{sA}} \exp\left(-\frac{(x_A - x_{Ap})^2}{2\sigma_{sA}^2}\right), \tag{2.4}$$

where σ_{sV} and σ_{sA} are the standard deviations of visual and auditory noises and also correspond to visual and auditory spatial resolutions. Table 1 shows all the parameters used in this letter and their values. In order to reproduce experimental results on the ventriloquism aftereffect, the noise distributions of equations 2.3 and 2.4 are subject to an adaptive update at each stimulus presentation. A detailed explanation of our model of adaptation is provided later.

We define $P(x_{Vp}, x_{Ap}, \xi)$ in equation 2.1 as a joint prior distribution of x_{Vp} , x_{Ap} , and ξ , and this can be written as

$$P(x_{Vp}, x_{Ap}, \xi) = P(x_{Vp}, x_{Ap}|\xi)P(\xi). \tag{2.5}$$

The locations of audiovisual stimuli should be close if they originate from the same source; otherwise, they are not correlated. Further, we assume that the stimuli sources are distributed uniformly. We model $P(x_{Vp}, x_{Ap}|\xi)$ as a gaussian distribution, which is a function of the difference between the visual and auditory locations for $\xi = 1$, and as a constant distribution when $\xi = 0$, that is,

$$P(x_{Vp}, x_{Ap}|\xi) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_{sp}L_s} \exp\left(-\frac{(x_{Vp}-x_{Ap})^2}{2\sigma_{sp}^2}\right) & (\xi = 1), \\ \frac{1}{L_s^2} & (\xi = 0), \end{cases} \tag{2.6}$$

where L_s is the length of the spatial integral range for normalization and σ_{sp} is the standard deviation of the distance between audiovisual stimuli when they share a common source. We take $L_s \gg \sigma_{sp}$ so as to eliminate the boundary effect.

Table 1: List of Parameters and Their Values.

Parameter	Value1	Value2	Value3	Description
Spatial or temporal resolution				
σ_{sV}	2.5°	0.8°	0.8°	Spatial resolution of vision
σ_{sA}	8°	2.5°	2.5°	Spatial resolution of audition
σ_{sp}	1°	1°	1°	Standard deviation (SD) of the distribution of distance between audiovisual stimuli originating from the same source
σ_{tV}	—	45 ms	—	Temporal resolution of vision
σ_{tA}	—	10 ms	—	Temporal resolution of audition
σ_{tp}	—	10 ms	—	SD of the distribution of temporal separation between audiovisual stimuli originating from the same source
Spatial adaptation				
α_V	—	—	0.003	Coefficient used in visual adaptation
α_A	—	—	0.003	Coefficient used in auditory adaptation
Others				
W_s	8°	4°	4°	Spatial window width for the judgment of spatial unity
W_t	—	65 ms	—	Temporal window width for the judgment of simultaneity
L_s	180°	180°	180°	Spatial integral region
L_t	—	1000 ms	—	Temporal integral region
$P(\xi = 1)$	0.005	0.2	^a	The probability that an audiovisual stimulus pair originate from the same source

Note: The value1, value2, and value3 columns show the values used in simulation 1, simulation 2, and simulation 3, respectively, in section 3.

^aSee appendix B for value 3.

Putting all this together, equation 2.1 can be written as

$$P(x_{Vp}, x_{Ap}, \xi | x_V, x_A) = \frac{P(x_V | x_{Vp})P(x_A | x_{Ap})P(x_{Vp}, x_{Ap} | \xi)P(\xi)}{P(x_V, x_A)}, \quad (2.7)$$

which is the product of the visual and auditory sensory noise distributions, the prior distributions of visual and auditory stimuli positions, and the prior distribution of ξ , divided by a normalization factor.

Next, let us take temporal factors into consideration. The detailed calculation including temporal factors is provided in appendix A. The assumptions and calculation procedures are almost the same as the spatial factors. Assuming independent spatial and temporal noises and priors, we find that

$$P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi | x_V, x_A, t_V, t_A) = \frac{P(x_V | x_{Vp})P(x_A | x_{Ap})P(t_V | t_{Vp})P(t_A | t_{Ap})P(x_{Vp}, x_{Ap} | \xi)P(t_{Vp}, t_{Ap} | \xi)P(\xi)}{P(x_V, x_A, t_V, t_A)}. \quad (2.8)$$

Probability $P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi | x_V, x_A, t_V, t_A)$ can be decomposed into a product of noise terms and physical variation terms, each of which is related to either visual or auditory stimuli and to either space or time.

2.2 Modeling of Adaptation. We model the spatial adaptation of audiovisual perception by updating the distributions $P(x_V | x_{Vp})$ and $P(x_A | x_{Ap})$ of equations 2.3 and 2.4, respectively, at each stimulus presentation. We set two free parameters, μ_V and μ_A , to be adapted; this modifies the mean values of $P(x_V | x_{Vp})$ and $P(x_A | x_{Ap})$, respectively. Equations 2.3 and 2.4 are rewritten as

$$P(x_V | x_{Vp}) = \frac{1}{\sqrt{2\pi}\sigma_{sV}} \exp\left(-\frac{(x_V - x_{Vp} - \mu_V)^2}{2\sigma_{sV}^2}\right), \quad (2.9)$$

$$P(x_A | x_{Ap}) = \frac{1}{\sqrt{2\pi}\sigma_{sA}} \exp\left(-\frac{(x_A - x_{Ap} - \mu_A)^2}{2\sigma_{sA}^2}\right). \quad (2.10)$$

Each time the observer receives the adapting audiovisual stimuli, it estimates the corresponding parameters and updates the noise distributions based on its observations and estimations. The observer determines MAP estimators \hat{x}_{Vp} and \hat{x}_{Ap} from x_V and x_A and updates μ_V and μ_A as

$$\mu_V \rightarrow (1 - \alpha_V)\mu_V + \alpha_V(x_V - \hat{x}_{Vp}), \quad (2.11)$$

$$\mu_A \rightarrow (1 - \alpha_A)\mu_A + \alpha_A(x_A - \hat{x}_{Ap}), \quad (2.12)$$

where α_A and α_V represent the magnitude of adaptation for each adapting stimulus presentation.

It has been shown that the ventriloquism aftereffect is related to a coordinate transformation between different coordinates used by different

modalities (Lewald, 2002). We modeled this interpretation and shifted the mean values of $P(x_V|x_{Vp})$ and $P(x_A|x_{Ap})$ toward the biases at previous estimations. Such a shift leads to a shift in the visual and auditory reference frame.

3 Results

We consider the three experimental results listed below and perform computer simulations for each of them:

1. The relationship between the unity judgment and the ventriloquism effect
2. The spatial and temporal effect on audiovisual perception
3. The ventriloquism aftereffect

In the experiments we reproduce in this letter, the subjects were not instructed to report whether the audiovisual stimuli originated from the same source. Therefore, the posterior probability to be maximized is integrated with respect to ξ in the following sections.

3.1 MAP Estimation and the Ventriloquism Effect. In this section, we explain how our model causes the ventriloquism effect.

Due to sensory noises in auditory and visual sensory pathways, the perceived relative location of the sound with respect to the light, $x_A - x_V$, is distributed around the presented value $x_{Ap} - x_{Vp}$. Figure 1A shows the probability distribution,

$$P(x_A - x_V | x_{Ap} - x_{Vp}) = \iint_{x'_A - x'_V = x_A - x_V} P(x'_A | x_{Ap}) P(x'_V | x_{Vp}) dx'_A dx'_V, \quad (3.1)$$

for $x_{Ap} - x_{Vp} = 0^\circ$ and $x_{Ap} - x_{Vp} = 20^\circ$. In Figure 1B, we show the posterior distributions of $x_{Ap} - x_{Vp}$ given the perceived relative location $x_A - x_V$, that is,

$$P(x_{Ap} - x_{Vp} | x_A - x_V) = \sum_{\xi} \iint_{x'_{Ap} - x'_{Vp} = x_{Ap} - x_{Vp}} P(x'_{Ap}, x'_{Vp}, \xi | x_A, x_V) dx'_{Ap} dx'_{Vp} \quad (3.2)$$

for $x_A - x_V = 8^\circ$ (circles) and $x_A - x_V = 18^\circ$ (crosses). In Figure 1B, we observe that the posterior distributions have two peaks that are located near 0° and $x_A - x_V$, and the ratio of the magnitude of the peaks depends on $x_A - x_V$. When $x_A - x_V = 8^\circ$, the peak near 0° is higher; thus, the difference

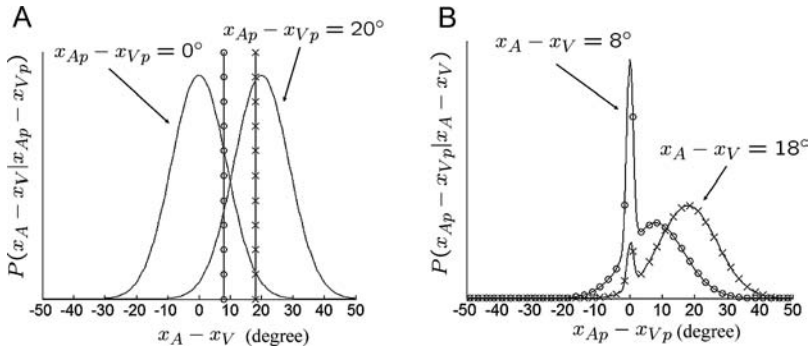


Figure 1: (A) The probability distribution of $x_A - x_V$ when the presented stimuli have $x_{Ap} - x_{Vp} = 0^\circ$ and 20° . (B) The posterior probability distribution of $x_{Ap} - x_{Vp}$ when the observed value is $x_A - x_V = 8^\circ$ (circles) and 18° (crosses). The lines with circles and crosses in A represent the cases in which $x_A - x_V$ is perceived to be 8° (circles) or 18° (crosses); each of them corresponds to the line with either circles or crosses in B.

of the estimations of x_{Ap} and x_{Vp} is near 0° , which implies that the ventriloquism effect has taken place. It should be noted that the peak around 0° mainly originates from $P(x_{Ap}, x_{Vp}, \xi = 1 | x_A, x_V)$ (the probability for an assumed common source) and the peak around $x_A - x_V$ originates from $P(x_{Ap}, x_{Vp}, \xi = 0 | x_A, x_V)$ (the probability for assumed uncommon sources).

3.2 Simulation 1: The Relationship Between the Spatial Unity Judgment and the Ventriloquism Effect. The purpose of simulation 1 is to confirm the results of the experiment by Wallace et al. (2004), who showed a strong relationship between the spatial unity perception (the perception that the light and sound come from the same location) and the ventriloquism effect. For simplicity, this section considers only the spatial factors.

Wallace et al. (2004) showed that when the spatial unity perception was reported, the localization of the auditory stimulus was almost biased to the location of the visual stimulus. However, when this perception was not reported, they observed a negative bias, that is, the localization of the auditory stimulus was rather biased opposite to the location of the visual stimulus. They also found that the standard deviation of auditory localization was relatively small when the spatial unity was reported compared to when it was not reported.

We performed a computer simulation to compare the consequences of our model with the experimental results. In simulation 1, the task was to localize the auditory stimuli and judge the spatial unity. Gaussian noises with mean 0° and standard deviations σ_{sV} and σ_{sA} were added to the presented stimuli x_{Ap} and x_{Vp} , which resulted in x_V and x_A ,

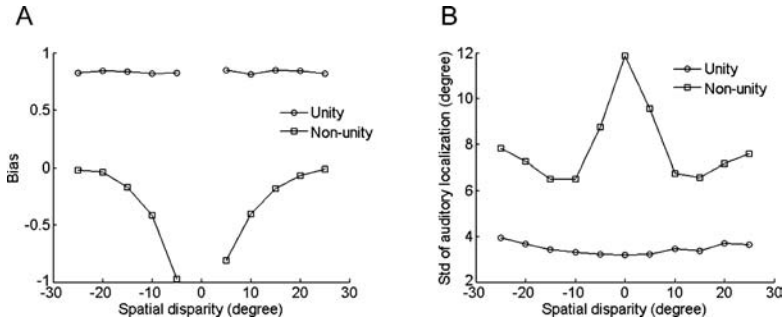


Figure 2: (A) The relation between the strength of the ventriloquism effect and the judgment of the spatial unity. The vertical axis represents the auditory localization bias. (B) Standard deviation of the auditory localization. The vertical axis represents the standard deviation of the auditory localization. In both the figures, the horizontal axis represents the spatial disparity between the light and the sound ($x_{Vp} - x_{Ap}$), and the two lines correspond to unity ($|\hat{x}_{Vp} - \hat{x}_{Ap}| < W_s$) and nonunity ($|\hat{x}_{Vp} - \hat{x}_{Ap}| > W_s$) cases.

respectively. The observer determined the estimations \hat{x}_{Vp} and \hat{x}_{Ap} by maximizing $\sum_{\xi} P(x_{Vp}, x_{Ap}, \xi | x_V, x_A)$ and then judged the spatial unity as yes if $|\hat{x}_{Vp} - \hat{x}_{Ap}| < W_s$. It should be noted that the spatial unity perception ($|\hat{x}_{Vp} - \hat{x}_{Ap}| < W_s$) and the perception of the common cause (MAP estimator of ξ equals 1) are different. If audiovisual stimuli are presented at the same position but with great temporal disparity, the stimuli are unlikely to have originated from the same event; however, subjects mostly report spatial unity because the stimuli are presented at the same place.

The strength of the auditory localization bias was defined as

$$\text{bias} = \frac{\hat{x}_{Ap} - x_{Ap}}{x_{Vp} - x_{Ap}}, \quad (3.3)$$

where bias = 1 implies that the auditory localization was completely attracted to the visual stimulus, and a minus bias implies that it was repulsed from the visual location.

The trial was repeated 2000 times for each audiovisual stimulus pair. Other details of the simulation are provided in appendix B. After the calculation, the results were classified into two groups depending on the spatial unity judgment.

Figure 2A shows that when the spatial unity was judged, the auditory localization was strongly attracted by the visual signal. On the contrary, when it was not judged, the bias was rather negative, and the auditory localization was biased opposite to the visual stimulus (Wallace et al., 2004). Further, this negative bias was greater when the stimuli were presented in

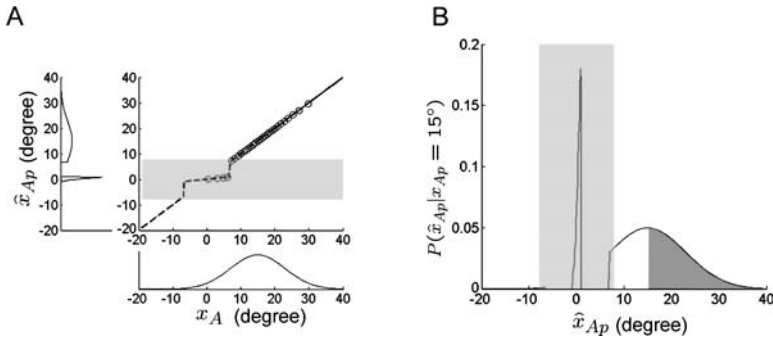


Figure 3: (A) Upper-right panel: relation of \hat{x}_{Ap} to x_A . Each circle is located at (x_A, \hat{x}_{Ap}) , where x_A is generated according to the probability distribution of x_A when $x_{Vp} = x_V = 0^\circ$, $x_{Ap} = 15^\circ$ and \hat{x}_{Ap} is the corresponding estimation. The shaded area represents the unity case. The horizontal axis represents x_A , and the vertical axis represents \hat{x}_{Ap} . The bottom and left panels represent the probability distributions of x_A and \hat{x}_{Ap} , respectively, when $x_{Ap} = 15^\circ$. (B) The probability distribution of \hat{x}_{Ap} when $x_{Ap} = 15^\circ$. The brighter-shaded area represents the unity case. The darker-shaded area represents cases where the bias is negative. The horizontal axis represents \hat{x}_{Ap} , and the vertical axis represents the probability density of \hat{x}_{Ap} .

close proximity. Figure 2B shows the standard deviation of the auditory localization for each test stimulus. From Figure 2B, we observe that the standard deviation of the auditory localization was greater in the nonunity case, where the localization deviation increased as the spatial disparity decreased (Wallace et al., 2004). These results agree with the experimental results by Wallace et al.

Figure 3 explains the reason for the negative bias in the nonunity case in our model. In this figure, we assume $x_{Vp} = x_V = 0^\circ$ for simplicity. The upper-right panel of Figure 3A shows the relation of \hat{x}_{Ap} to x_A . Each circle is located at (x_A, \hat{x}_{Ap}) , where x_A is generated according to equation 2.4 when $x_{Ap} = 15^\circ$ and \hat{x}_{Ap} is the corresponding MAP estimator. The spatial unity is judged as yes in the shaded area ($|\hat{x}_{Ap}| < W_s$). The bottom and left panels of Figure 3A show the probability distributions of x_A and \hat{x}_{Ap} , respectively, when $x_{Ap} = 15^\circ$. Figure 3B again shows the probability distribution of \hat{x}_{Ap} when $x_{Ap} = 15^\circ$. The spatial unity is judged as yes in the brighter-shaded area ($|\hat{x}_{Ap}| < W_s$). In the darker-shaded area ($\hat{x}_{Ap} > x_{Ap}$), the estimated location is greater than the presented location, which means the bias is negative. Due to the relatively large value of W_s , the sharp distribution near 0° is always included in the brighter-shaded area. Therefore, it can be observed that the mean bias becomes negative in the nonunity case (outside the brighter-shaded area).

3.3 Simulation 2: Spatial and Temporal Effect on Audiovisual Perception. In simulation 2, we investigated the temporal dependency of the spatial unity judgment and the spatial dependency of the simultaneous judgment. Both spatial and temporal factors were considered in this simulation.

The temporal discrepancy between visual and auditory stimuli influenced the perception of the spatial unity (Slutsky & Recanzone, 2001). It was found that the proportion of the spatial unity judgment decreased as the temporal discrepancy increased. It was also shown that the spatial discrepancy modulates the perception of simultaneity or time resolution of the audiovisual stimuli (Bertelson & Aschersleben, 2003; Lewald & Guski, 2003; Zampini et al., 2005). One possible explanation for the temporal dependency of the spatial unity is as follows. As described above, when the temporal disparity is large, the ventriloquism effect is small. Therefore, audiovisual stimuli are perceived to be more spatially distant in this case than when the temporal disparity is small; thus, the stimuli are likely to be perceived as originating from different locations.

Let us consider the task of reporting the spatial unity in our model. When the temporal discrepancy is large, the audiovisual stimuli are not likely to originate from the same source, and the ventriloquism effect is absent. This leads to a decrease in the spatial unity perception. In mathematical terms, by omitting normalization factors in the denominators in equations 2.7 and 2.8, we observe that $P(t_V|t_{Vp})P(t_A|t_{Ap})P(t_{Vp}, t_{Ap}|\xi)P(\xi)$ in equation 2.8 corresponds to $P(\xi)$ in equation 2.7. Therefore, the decision of the spatial unity depends on the temporal disparity via ξ .

We performed two computer simulations that investigated different tasks for audiovisual stimuli with various spatial and temporal disparities: Simulation 2.1 for the spatial unity judgment task and simulation 2.2 for the simultaneity judgment task. In simulation 2.1, since the observer did not judge the temporal properties, it determined estimators \hat{x}_{Vp} and \hat{x}_{Ap} by maximizing the posterior probability,

$$P(x_{Vp}, x_{Ap}|x_V, x_A, t_V, t_A) = \iint \sum_{\xi} P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi|x_V, x_A, t_V, t_A) dt_{Vp} dt_{Ap}. \quad (3.4)$$

Then the observer judged the spatial unity as yes if $|\hat{x}_{Vp} - \hat{x}_{Ap}| < W_s$. Similarly, in simulation 2.2, the observer determined estimators \hat{t}_{Vp} and \hat{t}_{Ap} by maximizing

$$P(t_{Vp}, t_{Ap}|x_V, x_A, t_V, t_A) = \iint \sum_{\xi} P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi|x_V, x_A, t_V, t_A) dx_{Vp} dx_{Ap}, \quad (3.5)$$

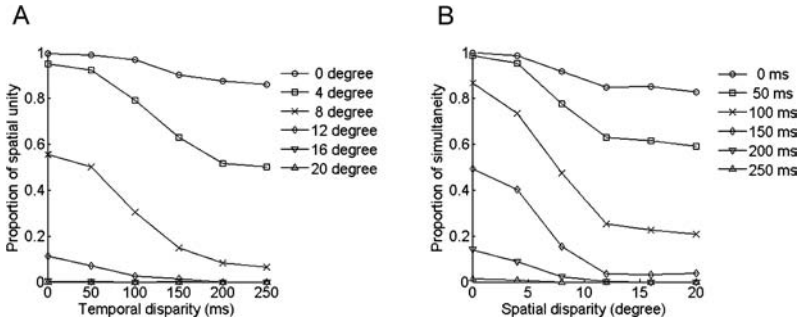


Figure 4: (A) Temporal dependency of the spatial unity judgment. The vertical axis represents the proportion of the spatial unity judgment. The horizontal axis denotes the temporal disparity, $t_{Vp} - t_{Ap}$, between the audiovisual stimuli. Each line represents the data calculated for the different spatial disparity, $x_{Vp} - x_{Ap}$. (B) Spatial dependency of the simultaneity judgment. The vertical axis represents the proportion of the simultaneity judgment. The horizontal axis denotes the spatial disparity between audiovisual stimuli. The different lines were calculated for different temporal disparities.

and judged the temporal simultaneity as yes if $|\hat{t}_{Vp} - \hat{t}_{Ap}| < W_t$. Other details of the simulations are provided in appendix B.

The simulation results for simulations 2.1 and 2.2 are shown in Figures 4A and 4B, respectively. In Figure 4A, we observe that the proportion of the spatial unity judgment decreases as the temporal disparity increases. This temporal dependency of the spatial unity judgment is consistent with the experimental results (Slutsky & Recanzone, 2001). Further, Figure 4B shows that the proportion of the simultaneity judgment decreases as the spatial disparity increases. This result is also consistent with the experimental results (Zampini et al., 2005).

3.4 Simulation 3: The Ventriloquism Aftereffect. In simulation 3, we investigated the audiovisual spatial adaptation known as the ventriloquism aftereffect. In this simulation, only spatial factors were considered.

It is known that after a prolonged exposure to simultaneous audiovisual stimuli with spatial disparity, the disparity that strongly produces a spatial unity perception shifts toward the presented disparity (Recanzone, 1998). The aftereffect can also be observed in the unimodal auditory localization.

First, we presented the adapting stimuli and performed an adaptation process. The light and the sound were presented at the same relative positions; the light was 8° to the right of the sound. The adapted parameters μ_V and μ_A were initially set to zero and were updated after each presentation of the adapting stimuli. The adapting stimuli were presented 2500 times. Test audiovisual stimuli were presented after the completion of the

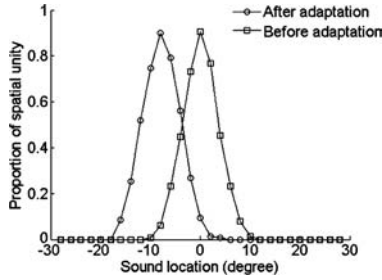


Figure 5: Results of spatial adaptation for audiovisual test stimuli after the 8° adaptation condition. The vertical axis represents the proportion of the spatial unity judgment. The horizontal axis denotes the location of the presented auditory stimulus. The visual stimulus was fixed at 0° .

adaptation. The observer estimated \hat{x}_{Vp} and \hat{x}_{Ap} and judged the spatial unity. Other simulation details are provided in appendix B.

Figure 5 shows the result for audiovisual test stimuli after the presentation of the adapting stimuli with $x_{Vp} - x_{Ap} = 8^\circ$. This figure shows the proportion of judgment of the spatial unity for various auditory test locations when the visual stimulus was fixed at 0° .

Figure 5 indicates that after the adaptation, spatial unity was more likely to be judged when the light was located relatively to the right of the sound; this location was consistent with the relative position of the light during the adaptation period. This result agrees with the experimental results (Recanzone, 1998).

We also investigated the adaptation effect on the unimodal localization, that is, the localization when only the test auditory or visual stimulus was presented. We assumed that the unimodal location was determined by maximizing the posterior probability of the physical position. For the auditory localization, the observer received x_A and maximized $P(x_{Ap}|x_A)$. From Bayes' formula, this is given by

$$P(x_{Ap}|x_A) = \frac{P(x_A|x_{Ap})P(x_{Ap})}{P(x_A)}. \quad (3.6)$$

If we assume an isotropically distributed stimulus source and a uniform prior distribution $P(x_{Ap})$, the estimator \hat{x}_{Ap} is the one that maximizes $P(x_A|x_{Ap})$. Other simulation details are provided in appendix B.

Figure 6 shows the results for the visual and auditory unimodal test stimuli. The simulation details are provided in appendix B. The auditory unimodal localization was shifted in the relative direction of the light in the adaptation period. However, the visual unimodal localization was not greatly affected by the adaptation. This relatively large shift in the

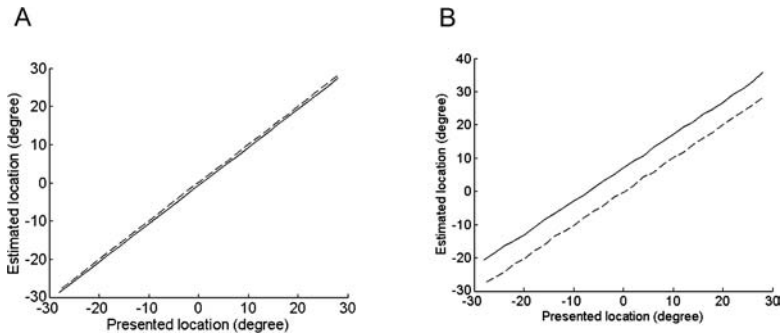


Figure 6: Unimodal (A) visual and (B) auditory localization after adaptation. The solid line represents the estimated locations for the various presented locations after adaptation. The broken line shows the estimation before adaptation.

auditory localization and the small shift in the visual localization agree with the experimental results of the ventriloquism aftereffect (Lewald, 2002; Recanzone, 1998).

The relatively great shift in the auditory unimodal localization is due to our assumption that the adaptation was executed based on estimated parameters. The estimation of the auditory location differed from the perceived location during the adaptation period because of the ventriloquism effect. However, the estimation of the visual location was not distant from the perceived location. This is why in our model, the auditory localization was strongly affected by the adaptation, but the visual localization was not greatly affected by it.

4 Summary and Discussion

In this letter, we modeled audiovisual integration using adaptive Bayesian inference. We defined a Bayesian observer that uses Bayesian inference to judge the locations or the timing of stimuli in addition to whether audiovisual stimuli are from the same source. The idea that the source of the audiovisual stimuli plays a crucial role in audiovisual integration has been termed the unity assumption (Welch & Warren, 1980) or pairing (Epstein, 1975). While former models on audiovisual interactions with Bayesian inference (Battaglia et al., 2003; Witten & Knudsen, 2005) assumed that audiovisual stimuli are bound from the beginning, our model not only considers positions and the timing of audiovisual stimuli but also considers whether audiovisual stimuli should be bound. It also explains audiovisual adaptation by adaptively changing the probability distributions of sensory noises.

The observer estimates the physically delivered position or timing of stimuli from noisy information. Although we applied the MAP rule in this

letter, it is also possible to apply the Bayes estimator that corresponds to the expected value of the posterior probability. We could derive qualitatively similar results using this rule as well.

We showed that our model reproduces several experimental results on the binding of multisensory signals. Our model showed a strong relation between the spatial unity perception and the ventriloquism effect. We also showed that the audiovisual temporal discrepancy affects the perception of the spatial unity and that the spatial distance also affects the perception of simultaneity.

Further, we assumed that adaptive changes in noise distributions modify the audiovisual perception. We adapted the mean value of noise distributions based on the observed and estimated parameters, considering that the ventriloquism aftereffect may correspond to the adaptation of coordinate transformation. We showed that our model of adaptation reproduces the experimental results on spatial audiovisual adaptation known as the ventriloquism aftereffect.

Theoretically, there are two ways to include the adaptation in our model. One is to change the noise distributions $P(x_A|x_{Ap})$ and $P(x_V|x_{Vp})$, as shown in our model. These distributions characterize the noise property of the sensory system. Thus, the adaptation of these distributions is meant to keep track of the changes in one's inner sensory processing pathways with the aid of other sensory modalities. The other possibility is to change the prior distribution $P(x_{Vp}, x_{Ap}, \xi)$ in order to adapt to the changes in the stimulus properties of the external world (Sharpee et al., 2006; Smirnakis, Berry, Warland, Bialek, & Meister, 1997). The ventriloquism aftereffect is usually considered to be a recalibration process that compensates for the change in the relationship between physical input and inner representation, which is caused by development or injury to sensory organs (De Gelder & Bertelson, 2003). Further, it is difficult to imagine the situation where the physical direction of the light and the sound from the same event have a constant disparity during minutes or tens of minutes (this is the typical timescale needed to induce the ventriloquism aftereffect). However, the comparison of the audiovisual aftereffect caused by these two different adaptation mechanisms, that is, the adaptation to internal or external changes, is important; both of these adaptations might be plausible in reality.

Although our model can reproduce some aspects of audiovisual spatial adaptation, it cannot explain all the properties of the experimental results. For example, if we present the adapting stimuli with more spatial or temporal disparities than those in our simulation, the adaptation effect may disappear; however, some experimental results have suggested that the adaptation effect can be observed even when the disparities are relatively large (Fujisaki, Shimojo, Kashino, & Nishida, 2004; Lewald, 2002). Therefore, the adaptation mechanism can be partially, but not completely, described by our model. There may exist some other mechanisms for adaptation that have not been included in our adaptation model.

Our model has many parameters, as shown in Table 1. It is difficult to determine these parameter values because we cannot reproduce all the experiments with the same set of parameter values. However, we can predict the parameter dependency of experimental results from our model. For example, parameter L_t is the temporal integral region width, and this corresponds to the temporal range in which the stimuli are presented. The value of L_t influences the ratio of $P(t_{Vp}, t_{Ap}|\xi = 1)$ and $P(t_{Vp}, t_{Ap}|\xi = 0)$, as shown in appendix A. We can show that the proportion of temporal simultaneity judgment increases as L_t increases. In fact, it is experimentally known that when a wide range of temporal disparities was presented, the proportion of simultaneity judgment increased (Zampini et al., 2005), which is in agreement with our results.

Although we modeled spatial and temporal audiovisual integration, the mathematical formulation of our model is not limited to this integration. It is possible that the same formulation can be applied to the integration of other modalities such as visuotactile or sensorimotor integration. Our future work is to model such various multisensory integrations.

Appendix A: Bayesian Estimation with Spatial and Temporal Factors —

In appendix A, we introduce Bayesian inference, taking temporal factors into consideration. In addition to the notations described in section 2.1, let t_{Vp} and t_{Ap} represent times when visual and auditory stimuli, respectively, are presented to the observer, and t_V and t_A represent the noisy information that the observer can use to infer the actual physical times. Taking spatial and temporal factors into consideration, the observer determines the physical parameters based on conditional probability $P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi | x_V, x_A, t_V, t_A)$. From Bayes' formula, we obtain

$$P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi | x_V, x_A, t_V, t_A) = \frac{P(x_V, x_A, t_V, t_A | x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi) P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi)}{P(x_V, x_A, t_V, t_A)}. \quad (4.1)$$

We assume that the first term in the numerator on the right-hand side does not depend on ξ , as seen in section 2.1, and that there is no correlation between the noises added during spatial information processing and temporal information processing. Further, we assume that visual and auditory noises are independent in both spatial and temporal dimensions. It then follows that

$$\begin{aligned} P(x_V, x_A, t_V, t_A | x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi) &= P(x_V, x_A, t_V, t_A | x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}) \\ &= P(x_V, x_A | x_{Vp}, x_{Ap}) P(t_V, t_A | t_{Vp}, t_{Ap}) \\ &= P(x_V | x_{Vp}) P(x_A | x_{Ap}) P(t_V | t_{Vp}) P(t_A | t_{Ap}). \end{aligned} \quad (4.2)$$

With regard to the prior distribution term $P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi)$, we assume that the spatial and temporal physical variations of visual and auditory stimuli are independent. The term can then be written as

$$\begin{aligned} P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi) &= P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}|\xi)P(\xi) \\ &= P(x_{Vp}, x_{Ap}|\xi)P(t_{Vp}, t_{Ap}|\xi)P(\xi). \end{aligned} \quad (4.3)$$

Taking all the assumptions together, equation 4.1 is decomposed into

$$\begin{aligned} &P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi|x_V, x_A, t_V, t_A) = \\ &\frac{P(x_V|x_{Vp})P(x_A|x_{Ap})P(t_V|t_{Vp})P(t_A|t_{Ap})P(x_{Vp}, x_{Ap}|\xi)P(t_{Vp}, t_{Ap}|\xi)P(\xi)}{P(x_V, x_A, t_V, t_A)}. \end{aligned} \quad (4.4)$$

The distributions regarding the temporal factors are set the same as the spatial factors as follows:

$$P(t_V|t_{Vp}) = \frac{1}{\sqrt{2\pi}\sigma_{tV}} \exp\left(-\frac{(t_V - t_{Vp})^2}{2\sigma_{tV}^2}\right), \quad (4.5)$$

$$P(t_A|t_{Ap}) = \frac{1}{\sqrt{2\pi}\sigma_{tA}} \exp\left(-\frac{(t_A - t_{Ap})^2}{2\sigma_{tA}^2}\right), \quad (4.6)$$

$$P(t_{Vp}, t_{Ap}|\xi) = \begin{cases} \frac{1}{\sqrt{2\pi}\sigma_{tp}L_t} \exp\left(-\frac{(t_{Vp} - t_{Ap})^2}{2\sigma_{tp}^2}\right) & (\xi = 1), \\ \frac{1}{L_t^2} & (\xi = 0), \end{cases} \quad (4.7)$$

where σ_{tV} and σ_{tA} are the standard deviations of visual and auditory noises, respectively, in temporal information processing. These parameters correspond to visual and auditory temporal resolutions, respectively. L_t is the length of the temporal integral range for normalization, and σ_{tp} is the standard deviation of the temporal disparity distribution of audiovisual stimuli when they share the same source.

Appendix B: Simulation Details

B.1 Simulation 1. The parameter values are shown in Table 1 (value1). The spatial axis was divided into 0.5° width bins, and the results were calculated numerically. Physical locations of auditory stimuli were fixed at $x_{Ap} = 0^\circ$, and a visual stimulus was delivered to one of $\{-25^\circ, -20^\circ, \dots, +25^\circ\}$. Gaussian noises with mean 0° and standard deviations σ_{sV} and σ_{sA} were

added to presented stimuli x_{Vp} and x_{Ap} , respectively, which resulted in x_V and x_A . The observer determined estimations \hat{x}_{Vp} and \hat{x}_{Ap} by maximizing $\sum_{\xi} P(x_{Vp}, x_{Ap}, \xi | x_V, x_A)$ and then judged the spatial unity as yes if $|\hat{x}_{Vp} - \hat{x}_{Ap}| < W_s$. Trials for each stimulus position were repeated 2000 times. The number of iterations was roughly set to follow the experiment by Wallace et al. (2004). Since the standard deviations σ_{sV} and σ_{sA} were not provided in the paper, we adopted the values that were measured in another experiment (Hairston et al., 2003), which was conducted by the same group.

B.2 Simulation 2. The parameter values used in simulation 2.1 are shown in Table 1 (value2). The spatial axis was divided into 0.3° width bins, and the temporal axis was divided into 5 ms bins. The physical position of the auditory stimulus was fixed at $x_{Ap} = 0^\circ$. The zero point of time was set to the physical arrival time of the auditory stimulus: $t_{Ap} = 0$ ms. Location x_{Vp} was one of $\{0^\circ, 4^\circ, 8^\circ, 12^\circ, 16^\circ, 20^\circ\}$, and t_{Vp} was one of $\{0$ ms, 50 ms, 100 ms, 150 ms, 200 ms, 250 ms $\}$. Trials were repeated 1000 times for each pair of visual and auditory stimuli. Each time the stimuli were presented, noises with zero mean and standard deviation σ_{sV} , σ_{sA} , σ_{tV} , and σ_{tA} were added to x_{Vp} , x_{Ap} , t_{Vp} , and t_{Ap} , respectively, which resulted in x_V , x_A , t_V , and t_A . The task was to judge spatial unity. The observer determined estimators \hat{x}_{Vp} and \hat{x}_{Ap} by maximizing $\int \int \sum_{\xi} P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi | x_V, x_A, t_V, t_A) dt_{Vp} dt_{Ap}$. The observer judged the spatial unity as yes if $|\hat{x}_{Vp} - \hat{x}_{Ap}| < W_s$. In simulation 2.2, all the parameter values and stimulus presenting procedures were the same as in simulation 2.1. The task was to judge simultaneity. In simulation 2.2, the observer determined estimators \hat{t}_{Vp} and \hat{t}_{Ap} by maximizing $\int \int \sum_{\xi} P(x_{Vp}, x_{Ap}, t_{Vp}, t_{Ap}, \xi | x_V, x_A, t_V, t_A) dx_{Vp} dx_{Ap}$ and judged the temporal simultaneity as yes if $|\hat{t}_{Vp} - \hat{t}_{Ap}| < W_t$.

In the experiments that we wished to reproduce in simulations 2 and 3, most of the parameters such as σ_{sA} and σ_{sV} were not given explicitly, and we could not always adopt the parameter values from other similar experiments. In fact, in the experiments by Slutsky and Recanzone (2001) and by Recanzone (1998), the spatial resolution is much better than that by Wallace et al. (2004), which we described in simulation 1. Therefore, we chose reasonable parameter values such that the simulation results could fit the experimental data.

B.3 Simulation 3. The parameter values are shown in Table 1 (value3). In simulation 3, $P(\xi = 1)$ was set to 0.6 during the adaptation period and to 0 during the test period to reproduce the experimental design by Recanzone (1998). The spatial axis was divided into 0.1° width bins. During the adaptation period, the position x_{Ap} of the auditory stimulus was chosen randomly from $\{-28^\circ, -24^\circ, \dots, +20^\circ\}$, and the visual stimulus was 8° to the right of the selected auditory stimulus position. The adapting stimuli were presented 2500 times. The details of the stimulus presenting procedures

and judgment procedures were the same as those in simulation 1 in both the adaptation and audiovisual test periods. During the audiovisual test period, the visual stimulus location was fixed at 0° , and the auditory stimulus location was one of $\{-28^\circ, -26^\circ, \dots, +28^\circ\}$; each stimulus was presented 150 times. During the unimodal test period, the presented test auditory or visual stimulus location was one of $\{-28^\circ, -26^\circ, \dots, +28^\circ\}$, and the estimated location was determined by maximizing equation 3.6. The test stimuli were presented 150 times for each stimulus location in both visual and auditory tests.

Acknowledgments

This research is partially supported by the Japan Society for the Promotion of Science, a Grant-in-Aid for JSPS fellows (18-06772), and a Grant-in-Aid for Scientific Research on Priority Areas—system study on higher-order brain functions—from the Ministry of Education, Culture, Sports, Science and Technology of Japan (17022012).

References

- Alais, D., & Burr, D. (2004). The ventriloquist effect results from near-optimal bimodal integration. *Current Biology*, *14*(3), 257–262.
- Battaglia, P. W., Jacobs, R. A., & Aslin, R. N. (2003). Bayesian integration of visual and auditory signals for spatial localization. *Journal of the Optical Society of America A—Optics Image Science and Vision*, *20*(7), 1391–1397.
- Bertelson, P., & Aschersleben, G. (2003). Temporal ventriloquism: Crossmodal interaction on the time dimension—1. Evidence from auditory-visual temporal order judgment. *International Journal of Psychophysiology*, *50*(1–2), 147–155.
- Bertelson, P., & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. *Perception and Psychophysics*, *29*(6), 578–584.
- Canon, L. K. (1970). Intermodality inconsistency of input and directed attention as determinants of nature of adaptation. *Journal of Experimental Psychology*, *84*(1), 141–147.
- De Gelder, B., & Bertelson, P. (2003). Multisensory integration, perception and ecological validity. *Trends in Cognitive Sciences*, *7*(10), 460–467.
- Deneve, S., Latham, P. E., & Pouget, A. (2001). Efficient computation and cue integration with noisy population codes. *Nature Neuroscience*, *4*(8), 826–831.
- Deneve, S., & Pouget, A. (2004). Bayesian multisensory integration and cross-modal spatial links. *Journal of Physiology—Paris*, *98*(1–3), 249–258.
- Epstein, W. (1975). Recalibration by pairing: A process of perceptual learning. *Perception*, *4*, 59–72.
- Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*, 429–433.
- Fujisaki, W., Shimojo, S., Kashino, M., & Nishida, S. (2004). Recalibration of audiovisual simultaneity. *Nature Neuroscience*, *7*(7), 773–778.

- Hairston, W. D., Wallace, M. T., Vaughan, J. W., Stein, B. E., Norris, J. L., & Schirillo, J. A. (2003). Visual localization ability influences cross-modal bias. *Journal of Cognitive Neuroscience*, *15*(1), 20–29.
- Howard, I. P., & Templeton, W. B. (1996). *Human spatial orientation*. New York: Wiley.
- Knill, D. C., & Pouget, A. (2004). The Bayesian brain: The role of uncertainty in neural coding and computation. *Trends in Neurosciences*, *27*(12), 712–719.
- Körding, K. P., & Wolpert, D. M. (2004). Bayesian integration in sensorimotor learning. *Nature*, *427*, 244–247.
- Lewald, J. (2002). Rapid adaptation to auditory-visual spatial disparity. *Learning and Memory*, *9*(5), 268–278.
- Lewald, J., & Guski, R. (2003). Cross-modal perceptual integration of spatially and temporally disparate auditory and visual stimuli. *Cognitive Brain Research*, *16*(3), 468–478.
- Morein-Zamir, S., Soto-Faraco, S., & Kingstone, A. (2003). Auditory capture of vision: Examining temporal ventriloquism. *Cognitive Brain Research*, *17*(1), 154–163.
- Pavani, F., Spence, C., & Driver, J. (2000). Visual capture of touch: Out-of-the-body experiences with rubber gloves. *Psychological Science*, *11*(5), 353–359.
- Radeau, M. (1994). Auditory-visual spatial interaction and modularity. *Cahiers de Psychologie Cognitive—Current Psychology of Cognition*, *13*(1), 3–51.
- Radeau, M., & Bertelson, P. (1974). Aftereffects of ventriloquism. *Quarterly Journal of Experimental Psychology*, *26*, 63–71.
- Rao, R. P. N. (2004). Bayesian computation in recurrent neural circuits. *Neural Computation*, *16*(1), 1–38.
- Recanzone, G. H. (1998). Rapidly induced auditory plasticity: The ventriloquism aftereffect. *Proceedings of the National Academy of Sciences of the United States of America*, *95*(3), 869–875.
- Sharpee, T. O., Sugihara, H., Kurgansky, A. V., Rebrik, S. P., Stryker, M. P., & Miller, K. D. (2006). Adaptive filtering enhances information transmission in visual cortex. *Nature*, *439*, 936–942.
- Slutsky, D. A., & Recanzone, G. H. (2001). Temporal and spatial dependency of the ventriloquism effect. *Neuroreport*, *12*(1), 7–10.
- Smirnakis, S. M., Berry, M. J., Warland, D. K., Bialek, W., & Meister, M. (1997). Adaptation of retinal processing to image contrast and spatial scale. *Nature*, *386*, 69–73.
- Wallace, M. T., Roberson, G. E., Hairston, W. D., Stein, B. E., Vaughan, J. W., & Schirillo, J. A. (2004). Unifying multisensory signals across time and space. *Experimental Brain Research*, *158*(2), 252–258.
- Welch, R. B., & Warren, D. H. (1980). Immediate perceptual response to intersensory discrepancy. *Psychological Bulletin*, *88*(3), 638–667.
- Witten, I. B., & Knudsen, E. I. (2005). Why seeing is believing: Merging auditory and visual worlds. *Neuron*, *48*(3), 489–496.
- Zampini, M., Guest, S., & Shore, D. I. (2005). Audio-visual simultaneity judgments. *Perception and Psychophysics*, *67*(3), 531–544.