Multimodal access to verbal name codes

MARIAN BERRYHILL

Dartmouth College, Hanover, New Hampshire
and University of Pennsylvania, Philadelphia, Pennsylvania

KESTUTIS KVERAGA

Dartmouth College, Hanover, New Hampshire
Massachusetts General Hospital, Boston, Massachusetts
and Harvard Medical School, Cambridge, Massachusetts

AND

LISA WEBB AND HOWARD C. HUGHES

Dartmouth College, Hanover, New Hampshire

Congruent information conveyed over different sensory modalities often facilitates a variety of cognitive processes, including speech perception (Sumbly & Pollack, 1954). Since auditory processing is substantially faster than visual processing, auditory–visual integration can occur over a surprisingly wide temporal window (Stein, 1998). We investigated the processing architecture mediating the integration of acoustic digit names with corresponding symbolic visual forms. The digits “1” or “2” were presented in auditory, visual, or bimodal format at several stimulus onset asynchronies (SOAs; 0, 75, 150, and 225 msec). The reaction times (RTs) for echoing unimodal auditory stimuli were approximately 100 msec faster than the RTs for naming their visual forms. Correspondingly, bimodal facilitation violated race model predictions, but only at SOA values greater than 75 msec. These results indicate that the acoustic and visual information *are* pooled prior to verbal response programming. However, full expression of this bimodal summation is dependent on the central coincidence of the visual and auditory inputs. These results are considered in the context of studies demonstrating multimodal activation of regions involved in speech production.

Environmental stimuli routinely produce multimodal sensory signals. These multimodal signals are initially encoded in separate sensory pathways that converge in certain subcortical (e.g., Dräger & Hubel, 1976; Jay & Sparks, 1984; Meredith & Stein, 1983) and cortical (e.g., Giard & Peronnet, 1999; Iacoboni, Woods, & Mazziotta, 1998; Kimura & Tamai, 1992; Wallace, Meredith, & Stein, 1992) areas. Behavioral studies have shown that congruent multimodal stimuli generally facilitate sensory processing (Todd, 1912), particularly when one signal is degraded or ambiguous (Bernstein, Chu, & Briggs, 1973; Sumbly & Pollack, 1954). The processing advantage conferred by presentation of two stimuli, relative to either stimulus presented alone, is called the *redundant signals effect* (Diederich, 1995; Diederich & Colonius, 1991; Diederich, Colonius, Bockhorst, & Tabelaing, 2003; J. Miller, 1982, 1986). A robust redundant signals effect is often observed in studies of multimodal processing. In these experiments, observers are presented with multimodal (auditory and visual [A + V]) or with unimodal (auditory [A] or visual [V]) stimuli, and are required to respond identically to all stimuli (e.g., manual simple reaction time

[RT], manual choice RT, or saccadic eye movement). The dependent variable in these behavioral paradigms can be either response accuracy (Ashby & Townsend, 1986; D. M. Green & Swets, 1966) or RT (Diederich & Colonius, 2004; Hughes, Nelson, & Aronchick, 1998; Hughes, Reuter-Lorenz, Nozawa, & Fendrich, 1994; J. Miller, 1982; Mordkoff & Yantis, 1991; Nickerson, 1973; Nozawa, Reuter-Lorenz, & Hughes, 1994; Raab, 1962; Townsend & Nozawa, 1995). The typical finding is that responses to multimodal signals are faster and more accurate than those to the unimodal signals, an effect that is also called *bimodal summation*.

There are several mechanisms that could produce the redundant signals effect, in general, and bimodal summation, in particular. Each of them represents a variant of parallel processing. To suggest that the processing of auditory and visual inputs occurs in parallel seems quite natural, since the initial processing of each modality occurs in different modality-specific pathways. *Parallel processing* simply means that information processing within these modality-specific pathways occurs concurrently. Processing may proceed in both the auditory and visual

H. C. Hughes, hch@dartmouth.edu

channels simultaneously, but need not end at the same time. Theoretical work has identified several important concepts related to parallel processing and has illustrated how they affect predicted levels of performance in a parallel processing architecture (e.g., Colonius, 1990; Grice, Canham, & Boroughs, 1984; J. Miller, 1982; Townsend & Ashby, 1983; Townsend & Nozawa, 1995; Townsend & Wenger, 2004). All of these theoretical developments assume that the time it takes to complete processing on each channel is a random variable. Given the stochastic nature of RTs, this seems a natural assumption (cf. Diederich, 1995; Townsend & Ashby, 1983). We make the additional assumption that the auditory stimuli activate the auditory but not the visual channel. We also assume the converse—that the visual stimuli activate the visual but not the auditory channel. This is known as the *assumption of selective influence* (see, e.g., Townsend & Nozawa, 1995).

One important factor that governs the time course of parallel processing is the operation that terminates processing. In redundant signals paradigms, processing may be terminated as soon as either channel completes its processing. In such cases, a separate decision rule is applied to each parallel channel, and the output of that decision rule is transmitted to a Boolean OR gate, such that the first channel to complete processing triggers a response. This general class of model has a long history in experimental psychology (Raab, 1962), and in the recent psychological literature has been called a *race model*, or *minimum completion time parallel processing* (cf. Townsend & Nozawa, 1995). Parallel processing in the minimum completion mode predicts that redundant signals will be processed more quickly than either component signal presented alone, if we assume that the time taken to complete processing on each channel is a random variable. The reason for this lies in the fact that the OR operator will trigger a response as soon as the first channel completes processing. Thus, the system always selects the minimum of two processing times, and it is known that the expected value of this minimum operator is less than or equal to the minimum of the expected values of the distribution of the individual channel processing times (e.g., Billingsley, 1979; Colonius & Diederich, 2006). This can be written as:

$$E[\min(T_A, T_V)] \leq \min[E(T_A), E(T_V)], \quad (1)$$

where T_A and T_V represent the processing times on the auditory and visual channels, respectively. Equation 1 is known as *Jensen's inequality*, and the facilitation it predicts is frequently called *probability summation*, because it is purely a statistical effect of applying the minimum completion time operator to a parallel processing architecture. Performance consistent with a race model architecture has been observed in a variety of visual tasks employing redundant signals, including the processing of noncorresponding binocular stimuli (Hughes & Townsend, 1998), motion detection (Meyer, Wuerger, Röhrbein, & Zetzsche, 2005; Wuerger, Hofbauer, & Meyer, 2003), and the concurrent processing of high and low spatial frequencies (Hughes, Nozawa, & Kitterle, 1996). Performance consistent with the race model architecture has also been observed in several studies of auditory–visual interactions, including audiovisual presentations to infants (Neil, Chee-Ruiter, Scheier, Lewkowicz,

& Shimojo, 2006) and saccades to spatially misaligned auditory and visual targets (Hughes et al., 1994). Manual RTs to bimodal targets are often consistent with a race model architecture, even if the targets are presented in a spatial register (Hughes et al., 1994).

However, it has been recognized for some time that parallel processing in the minimum completion time mode is not the sole processing architecture that can produce robust redundant signals effects: *Separate decisions followed by an OR gate* is not the only processing termination rule. In a seminal series of experiments, J. Miller (1982) proposed a parallel architecture in which the outputs of parallel channels could be combined before the application of a single criterion. When that single criterion was met, it would terminate processing and trigger a detection response. He termed this mechanism *coactive parallel processing*, and pointed out that the magnitude of the redundant signals effect in a coactive architecture could easily be greater than that produced by minimum completion time parallel processing. Drawing from basic theorems of probability theory, J. Miller (1982) introduced the idea that there should be an upper limit to the magnitude of statistical facilitation produced by parallel processing in the minimum completion mode. J. Miller (1982) reasoned that since the joint probability associated with observing two independent events A and B [$P(A \text{ and } B)$] is given by the product of the marginal probabilities [$P(A) \times P(B)$], the probability of observing either A or B is as follows:

$$P(A \text{ or } B) = P(A) + P(B) - [P(A) \times P(B)].$$

That is, $P(A \text{ or } B)$ equals the sum of the marginal probabilities minus their joint probability. We can reformulate the above expression to make predictions concerning RTs by expressing these probabilities as cumulative distributions of processing times. For instance, we can express a race model architecture in which the bimodal processing times for the visual and auditory channels are statistically independent, as follows:

$$P(RT_{AV} \leq t) \leq P(RT_A \leq t) + P(RT_V \leq t) - [P(RT_A \leq t) \times P(RT_V \leq t)], \quad (2)$$

where $P(RT_{AV} \leq t)$ is the probability that the processing on either the auditory channel or the visual channel is completed by time t , $P(RT_A \leq t)$ is the marginal probability that auditory processing is completed by time t , $P(RT_V \leq t)$ is the marginal probability that visual processing is completed by time t , and $[P(RT_A \leq t) \times P(RT_V \leq t)]$ is the joint probability that both auditory and visual processing is completed by time t . J. Miller (1982) noted that an upper bound on performance produced by a race model was provided by setting the joint probability term equal to 0. This is equivalent to assuming that the processing times on each channel demonstrate complete negative dependencies. Thus, J. Miller (1982) suggested that an upper (i.e., fast) bound of probability summation in minimum completion time parallel processing is given by

$$P(RT_{AV} \leq t) \leq [P(RT_A \leq t) + P(RT_V \leq t)]. \quad (3)$$

This expression is an instance of Boole's inequality. When applied to the performance limits of parallel race models,

it has frequently been called [J.] *Miller's inequality* or the *race model inequality* (RMI; cf. Townsend & Nozawa, 1995; Townsend & Wenger, 2004). Equation 3 provides a limit on the fastest performance attainable by a race model, given certain assumptions. We will discuss those assumptions at greater length in the next paragraph. In the case of multimodal inputs, the RMI states that the value of the bimodal cumulative distribution function (CDF) must be less than or equal to the sum of the unimodal CDFs for all times, t . RT performance that exceeds this boundary demonstrates a magnitude of multimodal facilitation that is incompatible with race models that assume noninteracting parallel channels (e.g., Diederich & Colonius, 1987; Gielen, Schmidt, & Van den Heuvel, 1983; Giray & Ulrich, 1993; Hughes et al., 1994; Krummenacher, Müller, & Heller, 2001; J. Miller, 1982, 1986; Patching & Quinlan, 2004). Violations of the RMI are observed in the latency of saccadic eye movements to spatially aligned bimodal targets (Colonius & Arndt, 2001; Frens, Van Opstal, & Van der Willigen, 1995; Hughes et al., 1998; Hughes et al., 1994), a finding that is consistent with an extensive anatomical and electrophysiological literature demonstrating that the superior colliculus, an important oculomotor center, is also a site of multimodal convergence—a site where individual neurons receive direct auditory and visual afferents (Dräger & Hubel, 1975a, 1975b; Meredith & Stein, 1983; Stein, Jiang, Wallace, & Stanford, 2001; Stein, Wallace, Stanford, & Jiang, 2002; recently reviewed in Holmes & Spence, 2005).

Several Constraints and Caveats

Formally, $P(RT_A \leq t)$ and $P(RT_V \leq t)$ are the marginal CDFs of processing times on the auditory and visual channels under conditions of bimodal stimulation. However, these marginal probabilities are not directly observable. In empirical studies, these cumulative probabilities are estimated from trials in which visual and auditory signals are presented in isolation. Using the distribution of unimodal RTs to estimate the CDFs of the marginal probabilities involves an assumption called *context independence* (cf. Colonius, 1990; Townsend & Nozawa, 1995; Townsend & Wenger, 2004). Context independence means that the activity within, say, an auditory input channel, is not altered in any way by the addition of a visual signal. It is evident, therefore, that use of the RMI assumes context independence for both the auditory and visual channels, because it would otherwise be inappropriate to use the unimodal trials to estimate the unimodal marginal probabilities on bimodal trials.

Recently, Townsend and Wenger (2004) have demonstrated how a third type of parallel processing architecture—an *interactive parallel architecture*—can produce RMI violations. In an interactive parallel architecture, activity within one input channel is transmitted to the other input channel. This sort of interaction between channels could easily be implemented by collateral connections between parallel channels, which are often observed in the neuroanatomy of afferent pathways. If the cross talk between channels is excitatory, the channels produce a positive interaction. If the cross talk is inhibitory, then the channels display negative interactions. Townsend and Wenger presented simulations that demonstrate that sufficiently

robust excitatory cross talk can produce violations of the RMI, even if processing is terminated by the minimum completion time stopping rule. Thus, positively dependent parallel processing in the minimum completion mode can mimic coactive parallel processing. In light of these recent theoretical developments, we can state that RMI violations rule out all race models that posit noninteractive parallel channels demonstrating context independence. By default, then, RMI violations imply either a coactive parallel architecture (also called a *channel summation architecture*) or a powerful, positively dependent interactive parallel architecture. The three distinguishable implementations of parallel processing are illustrated in Figure 1. As far as we are aware, there is no empirical method capable of unambiguously distinguishing between a coactive parallel architecture and a positively dependent parallel architecture. At present we must be content with the conclusion that empirical violations of the RMI rule out the architecture in Figure 1A, but are consistent with either of the architectures illustrated in Figures 1B and 1C.

It is also notable that, when we considered the effect of stochastic dependencies in the RMI, it appeared that positive dependencies slowed performance (relative to independent channels) and negative dependencies predicted faster RTs than the independent race model (Colonius, 1990). However, in an interactive parallel architecture, positive dependencies speed RTs, and negative dependencies slow RTs (Townsend & Wenger, 2004). Townsend and Wenger show that this discrepancy is due to the fact that interactive parallel processing violates the assumption of context independence. We cannot directly evaluate context independence empirically. This is why we cannot conclude which process (channel summation or channel interaction) is the source of any observed violations of the RMI.

The oculomotor system is one important sensorimotor subsystem that integrates visual and auditory inputs according to the channel summation processing architecture. However, it is entirely possible that other processing modules may also employ coactive parallel processing. In humans, speech and language are important cognitive systems that also integrate information from different sensory modalities. Speech sounds are often accompanied by the visual inputs arising from articulatory movements, and the combination of these auditory and visual inputs provides meaningful information to the listener (Dodd, 1977). The interaction between auditory and visual input channels is quite compellingly demonstrated by the McGurk effect, in which conflicting auditory, “ba,” and visual, “ga,” stimuli produce the percept of an intermediate, fused syllable, “da” (McGurk & MacDonald, 1976). This fusion of auditory and visual inputs is so robust that it occurs even in cases in which the gender of the voice and the face are incongruent (K. P. Green, Stevens, Kuhl, & Meltzoff, 1990). This illusion may be caused by inaccurate channel summation or by a process of conflict resolution. Subjective reports of phoneme perception are clearly insufficient to determine the answer.

Integration of auditory and visual inputs during speech processing has also been investigated using neuroimaging and electrophysiological techniques (Beauchamp, Argall, Bodurka, Duyn, & Martin, 2004; Calvert, 2001; Cal-

Figure 1. Three parallel processing architectures. Panel A represents independent parallel channels and the OR decision node (the race model architecture). Panel B represents coactive parallel channels and a summation decision node (coactive architecture). Panel C represents interactive parallel channels and an OR decision node (minimum completion time stopping rule).

vert, Campbell, & Brammer, 2000; Eimer, 2001; Eimer & Driver, 2000; Macaluso, George, Dolan, Spence, & Driver, 2004). However, enhanced blood oxygenation level differences (BOLD) signals in fMRI are also insufficient to distinguish whether processing is consistent with a parallel race model or a channel summation architecture. A channel summation architecture implies facilitatory convergence at the level of individual neurons, but the relatively poor spatial resolution of fMRI is not capable of determining whether enhanced BOLD responses to multimodal inputs are the result of multimodal convergence at the cellular level; the same effects could occur if distinct unimodal neural populations were closely interdigitated, as is known to sometimes occur in nonhuman cortical areas (Stein & Meredith, 1993). This confound has been well controlled by considering only activations that are superadditive ($\text{multimodal}_{1\&2} > \text{unimodal}_1 + \text{unimodal}_2$; Laurienti, Perrault, Stanford, Wallace, & Stein, 2005). But the correspondence between BOLD responses and neural activity is not clearly understood, and even if it were, issues relating to processing architecture must ultimately be resolved by studies of behavioral performance. Thus, appropriate behavioral demonstrations of the channel architecture are needed to explain specific cases of multimodal integration.

The present experiment examines the functional architecture underlying bimodal integration when corresponding signals arise from both the visual and auditory modalities, rather than from conflicting signals, as in the McGurk effect. The task employed is a simple verbal naming task; the experiment extends recent findings described by Berryhill, Kveraga, Webb, and Hughes (2005). In the unimodal conditions (A, V), the subjects were presented with Arabic digits or they heard the digit names and were instructed to verbally name the digit as quickly as possible (e.g., Berryhill et al., 2005; Davis, Moray, & Treisman, 1961; Fitts, 1964; Mowbray, 1960; Mowbray & Rhoades, 1959; Regan, 1981; Theios, 1975). In the bimodal condition (A + V), we presented the subjects with both the visual symbol and the corresponding auditory name, and they were required to vocalize the digit name. In the bimodal condition, we also included a stimulus onset asynchrony (SOA) manipulation, in an attempt to compensate for differences in temporal processing between visual and auditory modalities, because previous studies have shown the importance of central simultaneity of auditory–visual integration rather than physical simultaneity (Diederich & Colonius, 2004; Frens et al., 1995; Hughes et al., 1994). It is known that there is an extended temporal window over which these multisensory signals influence each other (Dixon & Spitz, 1980; Meredith & Stein, 1983, 1986a, 1986b, 1996; Munhall, Gribble, Sacco, & Ward, 1996; Stein & Meredith, 1993). In order to determine whether or not the results conformed to race or channel summation architecture, we calculated the RMI for four different SOA values spanning more than the difference observed between unimodal processing times (as estimated by the difference in unimodal RTs).

METHOD

Subjects

Seven undergraduates (3 male, 18–21 years old) participated. All reported having normal hearing and vision, and all were native speakers of English. They were compensated for their time by receiving additional course credit for experimental participation or by earning \$4/session (their choice). The experimental protocol was approved by the Dartmouth Committee for the Protection of Human Subjects, and each subject signed an informed consent document.

Stimuli

The visual stimuli consisted of the integers “1” and “2.” The digits were white numbers (1.8°, 27 cd/m²) presented against a black background on a 19-in. CRT display. We controlled the 57-cm viewing distance by asking the subjects to lean against a forehead rest. We presented the stimuli for 100 msec (six raster scans at 16.67 msec/scan). Stimulus presentation and timing were controlled by a PC running DOS, which yoked stimulus presentation to the vertical refresh rate of the CRT monitor. This stimulus presentation routine also set a bit in the machine’s parallel port synchronously with the first video raster frame containing stimuli, to determine RTs.

The auditory stimuli consisted of the set of spoken digit names “1” and “2.” The auditory stimuli were digitized recordings of a female voice speaking the digit names; they were presented using a standard PC sound card and computer speaker system.

The bimodal stimuli consisted of *congruent* visual and auditory stimuli, presented with the auditory stimulus trailing the visual stimulus by the following SOA values: 0, 75, 150, and 225 msec. At these SOA values, the asynchronies were not noticed by most subjects.

Response Recording

We recorded verbal responses using an audio microphone that amplified the signal and controlled a voice-activated switch. An experimenter monitored the audio responses for accuracy. A second computer digitized signals from the parallel port of the stimulus-presentation machine as well as the output of the voice-controlled switch at the sampling rate of 1000 Hz, thereby enabling millisecond accuracy of the vocal RTs. The microphone was highly directional, and the voice-activated switch was never triggered by the auditory stimuli, because the speakers were placed behind the microphone. We also adjusted the sensitivity of the microphone to ensure that only the subject’s voice could cause the voice-activated switch to trigger. False triggering would also have registered as RTs close to 0 msec for the 0-msec SOA and the auditory-alone trials. This false triggering was never observed.

Experimental Procedures

Experimental sessions consisted of five blocks of 100 trials each. Trials began with the placement of a fixation cross in the center of the CRT display for a 500–1,500-msec foreperiod selected from a rectangular distribution. The subjects were then presented with the digits “1” or “2” as visual, auditory, or bimodal stimuli. Trial types and probabilities are summarized in Table 1. For the bimodal stimuli at SOAs greater than 0, the first stimulus was always the visual stimulus. Thus, there were six different types of trials: auditory (A), visual (V), and bimodal (AV) at four SOAs (AV₀, AV₇₅, AV₁₅₀, and AV₂₂₅). The SOA manipulation was designed to compensate for the differences in auditory and visual RTs. For SOAs > 0, RTs were timed from the beginning of the first (i.e., the visual) stimulus. The subjects were instructed to name the presented digit as quickly and as accurately as possible. Feedback providing their average RT and accuracy was presented after each block. Added monetary incentives were provided if they beat their practice session mean RTs and maintained above 95% accuracy.

RMI Calculation

All analyses were performed on the individual subject RT distributions. First, CDFs were calculated for the unimodal auditory and visual conditions. Second, the RMI boundaries for the different

Table 1
Probability Distribution of Trial Types: Auditory (A), Visual (V), or Bimodal (AV), With Several Stimulus Onset Asynchronies (SOAs), Led First by the Visual Stimulus

Trial Type	Relative Frequency
Auditory (A)	.10
Visual (V)	.10
Bimodal SOA 0 (AV_0)	.20
Bimodal SOA 75 (AV_{75})	.20
Bimodal SOA 150 (AV_{150})	.20
Bimodal SOA 225 (AV_{225})	.20

SOA conditions were determined by adjusting the unimodal auditory CDF by the value of the SOA. For example, at the 75-msec SOA, the RMI is created by adding 75 msec to the unimodal auditory CDF before adding it to the unimodal visual CDF. The sum of

this adjusted auditory CDF and the unimodal visual CDF formed the lower boundary of the race model. The lower RMI boundary reflects the fastest possible responses consistent with race model processing. The lower RMI boundary and the experimental CDFs are plotted in Figure 2. Positive values found after subtracting the lower RMI boundary from the experimental CDFs indicate violations of the model (Figure 3).

Calculation of Capacity Coefficients

Townsend and Nozawa (1995) introduced a formal definition of processing capacity, although the general notion of capacity has a long history in the psychological literature. In essence, processing capacity is concerned with the speed with which an item is processed as the number of concurrently processed items increases. Capacity is termed "unlimited" if the marginal probabilities for each item are invariant when the number of other items processed on other channels increases. The independent channels race model is an example of an unlimited

Figure 2. Comparison of the CDFs and the RMI from 1 example subject for the four different bimodal SOA values (0, 75, 150, 225 msec). The top four panels plot the observed CDF (dark line) with the calculated RMI boundary (light line). When the CDF is to the right of the RMI, there is no violation of the inequality. When the CDF is to the left of the RMI, the lower boundary is crossed and the RMI is violated, which is indicative of channel summation architecture. The lower left panel presents the subject's mean RTs for each condition following the graphing conventions in Figure 4. The bottom right panel plots the differences between the calculated RMI boundary and the obtained bimodal CDF presented in the top four panels. Positive values of this difference indicate violations of the inequality.

Figure 3. RMI values for all 7 subjects at each bimodal SOA condition. These values are calculated per person by subtracting the calculated RMI boundary from the calculated CDF. The solid lines represent the group means for each SOA condition.

capacity system. Any processing system that exceeds unlimited capacity is said to have supercapacity. Townsend and Nozawa derived the formal notion of the capacity coefficient, which is defined as

$$C_o(t) = \ln[S_{AV}(t)] / \ln[S_A(t)] + \ln[S_V(t)], \quad (4)$$

where $C_o(t)$ is the capacity coefficient for minimum completion time parallel processing at time t , $\ln[S_{AV}(t)]$ is the natural logarithm of the bimodal survivor function at time t , $\ln[S_A(t)]$ is the natural logarithm of the auditory survivor function at time t , and $\ln[S_V(t)]$ is the natural logarithm of the visual survivor function at time t . The survivor function is the complement of the CDF, $S(t) = 1 - \text{CDF}(t) = 1 - P(T \leq t) = P(T \geq t)$. A system that demonstrates unlimited capacity at time t should have a capacity coefficient = 1.0 at time t . A system that demonstrates supercapacity at time t should have a capacity coefficient > 1.0 at time t .

Individual capacity coefficients were calculated following the method of Townsend and Wenger (2004). For each subject and each SOA value, survivor functions were calculated by first subtracting the CDF from 1. The capacity coefficient was calculated by taking the natural log of the bimodal survivor function as the numerator and the sum of the natural logarithms of each unimodal survivor function. In order to account for the SOAs, we increased the auditory survivor function by the amount of the SOA value, just as we did in computing the RMI for SOAs > 0.

RESULTS

Response Accuracy

No subject performed below 99% correct for any stimulus condition.

Response Times

Grand mean RTs for all subjects are illustrated in Figure 4. The mean RTs for unimodal visual and auditory digits are illustrated on the left (open circle and filled square, respectively). There is a clear difference between the unimodal visual and auditory RTs. In addition, RTs

Figure 4. Grand mean RTs for all 7 subjects for unimodal auditory digits (filled square), unimodal visual digits (open circle), and for bimodal digits as a function of the four different SOAs (filled circles).

to bimodal trials vary substantially as a function of SOA. As expected, at 0-msec SOA, RTs were equivalent to unimodal auditory digits because the processing times to identify the auditory stimuli were substantially faster (by about 100 msec) than unimodal visual processing times. Bimodal RTs slowed with increasing SOAs (visual preceding auditory) until, at an SOA of 225 msec, bimodal RTs were nearly equivalent to unimodal visual RTs. This indicates that this range of SOAs effectively compensated for the overall difference between visual and auditory processing times. This is important, because the multimodal summation effects depend on the amount of overlap between the unimodal distributions. Figure 5 illustrates the CDFs for 1 representative subject for each of the six stimulus conditions.

The ordering of RTs that was apparent in the grand means is quite visible in the individual subject data. Figure 2 compares the boundary predicted by the race inequality measure with the obtained bimodal CDFs for 1 subject at each of the four visual-auditory SOAs.

The top four panels illustrate the obtained bimodal CDF for the four different SOA conditions. Also illustrated is the race model boundary for that SOA. It can be seen that at 0-msec SOA there were no violations of the RMI. There were essentially no violations at 75-msec SOAs either. In contrast, SOAs of 150 and 225 msec produced substantial violations. Probability summation cannot account for the facilitation produced by bimodal stimuli delivered at these SOAs. Rather, these results strongly suggest channel summation architecture. The data are presented in a more concise fashion in the bottom right panel of Figure 2, which illustrates the difference between the race model boundary and the obtained bimodal CDFs (predicted-obtained). Figure 3 is a scatterplot of the RMI values for all 7 subjects for the four SOA conditions. The heavy line is the group mean RMI as a function of time. Inspection of Fig-

Figure 5. Example CDFs for each trial type from 1 subject. The CDFs are ordered along the following pattern: unimodal auditory, 0-msec bimodal, 75-msec bimodal, 150-msec bimodal, 225-msec bimodal, unimodal visual. RMI limits were calculated for each bimodal condition by shifting the unimodal auditory CDF by the amount of the SOA value. The observed experimental CDFs were compared with the RMI boundary in order to determine any RMI violations.

ure 3 reveals that some RMI violations occurred in at least some subjects at each SOA examined. It is evident that massive violations occurred at the two longest SOAs (150 and 225 msec). This is indicated by the solid lines in each panel, which inscribe the average value of the difference between the obtained and predicted bimodal CDFs as a function of time.

We calculated the capacity coefficients separately for each subject at each SOA. Townsend and Nozawa (1995) provided a proof that an observed violation of the race model inequality at time t entails supercapacity at time t [$C_o(t) > 1.0$]; see also Townsend and Wenger (2004). Evidence of supercapacity was observed in all subjects for auditory-visual SOAs of 150 and 225 msec. At the SOA values of 0 and 75 msec, there were individual subjects who never showed evidence of supercapacity at any time, consistent with the general observation that they also did not violate the RMI at these SOAs. In order to quantify the relative level of supercapacity at each SOA value, we calculated the percent of 10-msec bins demonstrating supercapacity. We subjected these values to a repeated measures ANOVA with SOA as the repeated measure. There was a main effect of SOA [$F(3,18) = 6.7, p = .003, \eta_p^2 = .53$]. Post hoc pairwise comparisons revealed significantly greater supercapacity measurements for the 225-msec SOA condition than for the 150-msec SOA condition ($p = .037$). However, the variability across subjects was very high and the differences between the mean 0-msec SOA and the 75- and 225-msec SOAs was actually greater than the difference between the 150- and 225-msec SOAs; see Table 2.

DISCUSSION

It is clear that either the visual form or the acoustic sound of an alphanumeric character can rapidly and automatically activate the corresponding verbal name code in literate adults (Berryhill et al., 2005). The goal of the present study was to characterize the processing architecture underlying parallel auditory and visual integration of written symbols and their corresponding spoken names. We did this by using a divided attention paradigm, in which subjects had to verbalize digit names in response to (1) their visual form, (2) their auditory name, or (3) the combination of both the visual forms and the auditory names. In order to compensate for well-known differences in the speed of processing visual and auditory stimuli (Woodworth & Schlosberg, 1954), we combined these bimodal stimuli using four different SOAs. This approach follows previous experiments in which subjects made saccades to spatially localized visual and auditory targets (Hughes et al., 1998; Hughes et al., 1994). In the present data, substantial differences in auditory and visual processing times were observed. Furthermore, violations of the RMI were only apparent when the auditory component of the bimodal stimuli was delayed by an amount approximating the average difference in unimodal processing times. Thus, no consistent violations of the RMI (J. Miller, 1982) were observed at SOAs of 0 or 75 msec (visual stimulus leading), but substantial and consistent

Table 2

Trial Type	Bins Showing Supercapacity (%)	
	<i>M</i>	<i>SE</i>
Bimodal SOA 0 (AV_0)	31.7	10.7
Bimodal SOA 75 (AV_{75})	56.8	16.6
Bimodal SOA 150 (AV_{150})	45.9	13.0
Bimodal SOA 225 (AV_{225})	67.8	10.3

violations were observed at SOAs of 150 and 225 msec. We therefore conclude that bimodal information facilitates naming latencies to an extent that is incompatible with a parallel race processing architecture.

This conclusion is largely confirmed by analyzing the capacity coefficients, although there are several discrepancies we need to consider. Capacity coefficients greater than 1 indicate supercapacity, and should be observed whenever the RMI is violated. That prediction is largely confirmed in these data. However, the converse does not appear to hold—capacity coefficients greater than 1.0 were often observed at times when the RMI had not been violated. Almost every SOA value for every subject had at least one time bin with a capacity coefficient greater than 1, but there was tremendous variability between subjects. There is also a discrepancy between the two measures regarding supercapacity at the 0- and 75-msec SOA values. Although no RMI violations are observed at these SOA values, there are capacity coefficient values greater than 1 for 6 out of 7 subjects in these conditions. Thus, in the present data set, the RMI appears to be the more conservative measure of determining whether observed performance is incompatible with the race model architecture. One possible reason for the discrepancies may be that at least 300 data points per condition are suggested when calculating capacity coefficients, a condition our study did not satisfy. We therefore emphasize the cases for which there is congruency between the two measures, and conclude that these data provide psychophysical evidence indicating that auditory name and visual form information are either (1) another example of coactive parallel processing of auditory and visual information or (2) an example of strong, positively dependent parallel bimodal processing (see Figures 1B and 1C). It might seem unlikely that potent interactions occur between auditory and visual afferent channels, since there are no known connections between the auditory and visual systems, at least early in processing. However, significant cross talk could readily occur later in the pathways that mediate visual and auditory digit recognition, so we believe that the interactive parallel model is just as viable a candidate as the channel summation model.

An important feature of previous work is that the empirical tests of channel summation architecture provide psychophysical results that mesh nicely with known physiology (e.g., Hughes et al., 1998; Hughes et al., 1994; Hughes & Townsend, 1998). In the following paragraphs, we consider where the bimodal summation or interactions underlying the observed degree of intersensory facilitation of naming latencies might be occurring in the brain.

Location of Multimodal Convergence

The manner in which individual sensory properties are bound to one object remains a fundamental question in sensory processing. Converging data suggest that this process arises from processing networks in multiple subcortical and cortical structures. Correspondingly, multimodal convergence seems to occur at multiple levels of processing (for recent reviews, see Amedi, von Kriegstein, van Atteveldt, Beauchamp, & Naumer, 2005; Calvert & Thesen, 2004). Convergence within the oculomotor system seems to be appropriately regarded as occurring at the interface between sensory and motor processing. In this case, it is probably not useful to attempt to categorize the “locus” of summation as occurring at either a sensory or motor stage, since physiological studies have identified neurons within the superior colliculus that have multimodal receptive fields (Bell, Meredith, Van Opstal, & Munoz, 2005; Colonius & Diederich, 2004; Diederich & Colonius, 2004; Doubell, Skaliara, Baron, & King, 2003; Dräger & Hubel, 1975a, 1975b; Fort, Delpuech, Pernier, & Giard, 2002; Hughes et al., 1998; Joassin, Maurage, Bruyer, Crommelinck, & Campanella, 2004; L. M. Miller & D’Esposito, 2005; Schneider & Kastner, 2005; Sparks, 1986; Stein & Meredith, 1993; Toldi, Fehér, & Wolff, 1986; Wallace & Stein, 2001; reviewed in Isa & Sasaki, 2002; but see Populin & Yin, 2002) and also generate a premotor discharge prior to specific trajectories of saccadic eye movements (e.g., Jay & Sparks, 1984; Meredith & Stein, 1986a; Patton, Belkacem-Boussaid, & Anastasio, 2002; Stein, 1978; Stein & Arigbede, 1972).

The function of multimodal integration in oculomotor control clearly seems to promote foveation of environmental events, regardless of the sensory modality that encodes those events. The simple fact that the same response can be accessed by stimulation to different sensory modalities requires parallel routes to the response mechanisms. But it does not necessarily imply the channel summation architecture. For example, Hughes et al. (1994) demonstrated that the exact same bimodal stimuli produce a level of intersensory facilitation that is consistent with a race model architecture when manual RTs are the measure of performance, but a level of facilitation that exceeds the race model bound when saccade RTs are the dependent measure. Clearly, not all cases of intersensory facilitation require the channel summation architecture. In the present task, multimodal integration of congruent information could easily have been consistent with the race model. Alternatively, the processing could have components arranged in series if, for example, visual alphanumeric forms need to be translated into the auditory phonological representations before accessing the phonological motor programs. Multimodal integration of the sort demonstrated by the McGurk effect (McGurk & MacDonald, 1976) is likely to result from some sort of competition between perceptual alternatives, and might therefore be expected to result in processing times that are incompatible with (i.e., slower than) a race model architecture. Of course, these behavioral demonstrations of channel summation cannot specify the locus of the effect within the human nervous system. To address the anatomical locus

of this processing, we must consider relevant neuroimaging studies.

Neural pathways involved in mediating reading aloud and/or auditory involve activity in Broca's area, which is critical to initiating spoken language (for a meta-analysis of PET and MRI studies, see Turkeltaub, Eden, Jones, & Zeffiro, 2002). Neuroimaging work suggests that frontoparietal networks are important in mediating attentional contributions (Macaluso et al., 2004) to auditory-visual integration during bimodal presentation (Talsma & Woldorff, 2005) and speech processing (Saito et al., 2005). Recent investigations suggest that separate networks process the sensory (superior colliculus, insula, and intraparietal sulcus) and perceptual (Heschl's gyrus, superior temporal sulcus, middle intraparietal sulcus, and inferior frontal gyrus) components of multisensory integration (L. M. Miller & D'Esposito, 2005). These interactive networks are essential to language processing (Geschwind, 1965; Mesulam, 1998; reviewed in Cabeza & Nyberg, 2000). In addition, the important role of feedback as well as feedforward connections (Foxy & Schroeder, 2005) may explicate activity observed in early sensory regions (i.e., Calvert et al., 1997, 2000; Falchier, Clavagnier, Barone, & Kennedy, 2002; Giard & Peronnet, 1999; Heim, Opitz, Müller, & Friederici, 2003; Kayser, Petkov, Augath, & Logothetis, 2005; Raij, Uutela, & Hari, 2000; Santi, Servos, Vatikiotis-Bateson, Kuratate, & Munhall, 2003; but see Laurienti et al., 2002; Olson, Gatenby, & Gore, 2002; van Atteveldt, Formisano, Blomert, & Goebel, 2007). The channel summation architecture revealed in the present work presumably operates in numerous situations in which visual inputs facilitate processing of acoustic information (Calvert et al., 1999; Giard & Peronnet, 1999).

SOAs and Intersensory Priming

The present results present significant violations of the RMI at SOA values of 150 and 225 msec. These SOA values overcompensate by more than the 100-msec difference in auditory and visual unimodal RTs. This surprisingly large temporal window of bimodal influence is consistent with studies examining perceived synchrony. Although there are tremendous individual differences (L. M. Miller & D'Esposito, 2005), lags greater than 250 msec between the visual and audio signals are needed for humans to detect them (Dixon & Spitz, 1980). Thus, observing an effect at the higher SOA values may reflect the window of audio-visual integration.

One possible concern with regard to our paradigm and the use of SOAs involves the possibility of sensory priming. Priming is a well-known behavioral improvement in performance that occurs following the repetition of stimuli (reviewed in Schacter & Buckner, 1998). Because we observed such robust channel summation at the higher SOA values (150, 225 msec), it was suggested by one reviewer that the visual stimulus might prime the auditory stimulus, thus reducing RTs. Evidence from experimental approaches with high temporal fidelity, MEG and ERP, suggests that these effects do not begin until approximately

200–250 msec following presentation of the stimulus (Kim, Lee, Shin, Kwon, & Kim, 2006; Marinkovic et al., 2003; Nagy & Rugg, 1989). Although these paradigms are not precisely identical to the present study, they do evaluate the onset of perceptual priming. The onset times are close to the 225-msec SOA employed in the present paradigm, but the 150-msec SOA is considerably shorter. If priming were underlying our findings, we would not predict the observation of such dramatic RMI violations at the 150-msec SOA value.

Implications of Coactive and Interactive Architectures

Multimodal signals are important constituents of perception. It is likely that multimodal interactions follow a number of different processing architectures, depending on the particular cortical and subcortical pathways involved. Therefore, testing processing architectures using multiple paradigms, including unimodal and bimodal stimuli, is an important tool in distinguishing between processing pathways for various tasks. In humans, visual and auditory modalities both access speech control centers; people are able to read out loud following visual presentation of words and are also able to repeat what someone has just told them. Given that these parallel routes do exist, there are several prominent ways in which the channels might interact. Because we observe greater facilitation than would be expected, given a pure parallel race, we conclude that either channel summation or some form of highly interactive parallel system underlies verbal responses to redundant auditory and visual stimuli.

AUTHOR NOTE

The authors thank Michael Wenger for helpful comments. We also thank Jim Townsend for constructive discussion and Leslie Blaha for providing code to check our capacity coefficients. We also thank the reviewers for their careful attention. Correspondence concerning this article should be addressed to H. C. Hughes, Department of Psychological and Brain Sciences, Dartmouth College, 6207 Moore Hall, Hanover, NH 03755 (hch@dartmouth.edu).

REFERENCES

- AMEDI, A., VON KRIEGSTEIN, K., VAN ATTEVELDT, N. M., BEAUCHAMP, M. S., & NAUMER, M. J. (2005). Functional imaging of human cross-modal identification and object recognition. *Experimental Brain Research*, **166**, 559-571.
- ASHBY, F. G., & TOWNSEND, J. T. (1986). Varieties of perceptual independence. *Psychological Review*, **93**, 154-179.
- BEAUCHAMP, M. S., ARGALL, B. D., BODURKA, J., DUYN, J. H., & MARTIN, A. (2004). Unraveling multisensory integration: Patchy organization within human STS multisensory cortex. *Nature Neuroscience*, **7**, 1190-1192.
- BELL, A. H., MEREDITH, M. A., VAN OPSTAL, A. J., & MUNOZ, D. P. (2005). Crossmodal integration in the primate superior colliculus underlying the preparation and initiation of saccadic eye movements. *Journal of Neurophysiology*, **93**, 3659-3673.
- BERNSTEIN, I. H., CHU, P. K., & BRIGGS, P. (1973). Stimulus intensity and foreperiod effects in intersensory facilitation. *Quarterly Journal of Experimental Psychology*, **25**, 171-181.
- BERRYHILL, M. E., KVERAGA, K., WEBB, L., & HUGHES, H. C. (2005). Effect of uncertainty on the time course for selection of verbal name codes. *Perception & Psychophysics*, **67**, 1437-1445.
- BILLINGSLEY, P. (1979). *Probability and measure*. New York: Wiley.

- CABEZA, R., & NYBERG, L. (2000). Imaging cognition II: An empirical review of 275 PET and fMRI studies. *Journal of Cognitive Neuroscience*, **12**, 1-47.
- CALVERT, G. A. (2001). Crossmodal processing in the human brain: Insights from functional neuroimaging studies. *Cerebral Cortex*, **11**, 1110-1123.
- CALVERT, G. A., BRAMMER, M. J., BULLMORE, E. T., CAMPBELL, R., IVERSEN, S. D., & DAVID, A. S. (1999). Response amplification in sensory-specific cortices during crossmodal binding. *NeuroReport*, **10**, 2619-2623.
- CALVERT, G. A., BULLMORE, E. T., BRAMMER, M. J., CAMPBELL, R., WILLIAMS, S. C. R., MCGUIRE, P. K., ET AL. (1997). Activation of auditory cortex during silent lipreading. *Science*, **276**, 593-596.
- CALVERT, G. A., CAMPBELL, R., & BRAMMER, M. J. (2000). Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex. *Current Biology*, **10**, 649-657.
- CALVERT, G. A., & THESEN, T. (2004). Multisensory integration: Methodological approaches and emerging principles in the human brain. *Journal of Physiology-Paris*, **98**, 191-205.
- COLONIUS, H. (1990). Possibly dependent probability summation of reaction time. *Journal of Mathematical Psychology*, **34**, 253-275.
- COLONIUS, H., & ARNDT, P. (2001). A two-stage model for visual-auditory interaction in saccadic latencies. *Perception & Psychophysics*, **63**, 126-147.
- COLONIUS, H., & DIEDERICH, A. (2004). Why aren't all deep superior colliculus neurons multisensory? A Bayes' ratio analysis. *Cognitive, Affective, & Behavioral Neuroscience*, **4**, 344-353.
- COLONIUS, H., & DIEDERICH, A. (2006). The race model inequality: Interpreting a geometric measure of the amount of violation. *Psychological Review*, **113**, 148-154.
- DAVIS, R., MORAY, N., & TREISMAN, A. (1961). Imitative responses and the rate of gain of information. *Quarterly Journal of Experimental Psychology*, **13**, 78-89.
- DIEDERICH, A. (1995). Intersensory facilitation of reaction time: Evaluation of counter and diffusion coactivation models. *Journal of Mathematical Psychology*, **39**, 197-215.
- DIEDERICH, A., & COLONIUS, H. (1987). Intersensory facilitation in the motor component? A reaction time analysis. *Psychological Research*, **49**, 23-29.
- DIEDERICH, A., & COLONIUS, H. (1991). A further test of the superposition model for the redundant-signals effect in bimodal detection. *Perception & Psychophysics*, **50**, 83-86.
- DIEDERICH, A., & COLONIUS, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, **66**, 1388-1404.
- DIEDERICH, A., COLONIUS, H., BOCKHORST, D., & TABELING, S. (2003). Visual-tactile interaction in saccade generation. *Experimental Brain Research*, **148**, 328-337.
- DIXON, N. F., & SPITZ, L. (1980). The detection of auditory visual desynchrony. *Perception*, **9**, 719-721.
- DODD, B. (1977). The role of vision in the perception of speech. *Perception*, **6**, 31-40.
- DOUBELL, T. P., SKALIORA, I., BARON, J., & KING, A. J. (2003). Functional connectivity between the superficial and deeper layers of the superior colliculus: An anatomical substrate for sensorimotor integration. *Journal of Neuroscience*, **23**, 6596-6607.
- DRÄGER, U. C., & HUBEL, D. H. (1975a). Physiology of visual cells in mouse superior colliculus and correlation with somatosensory and auditory input. *Nature*, **253**, 203-204.
- DRÄGER, U. C., & HUBEL, D. H. (1975b). Responses to visual stimulation and relationship between visual, auditory, and somatosensory inputs in mouse superior colliculus. *Journal of Neurophysiology*, **38**, 690-713.
- DRÄGER, U. C., & HUBEL, D. H. (1976). Topography of visual and somatosensory projections to mouse superior colliculus. *Journal of Neurophysiology*, **39**, 91-101.
- EIMER, M. (2001). Crossmodal links in spatial attention between vision, audition, and touch: Evidence from event-related brain potentials. *Neuropsychologia*, **39**, 1292-1303.
- EIMER, M., & DRIVER, J. (2000). An event-related brain potential study of cross-modal links in spatial attention between vision and touch. *Psychophysiology*, **37**, 697-705.
- FALCHIER, A., CLAVAGNIER, S., BARONE, P., & KENNEDY, H. (2002). Anatomical evidence of multimodal integration in primate striate cortex. *Journal of Neuroscience*, **22**, 5749-5759.
- FITTS, P. M. (1964). Perceptual-motor skills learning. In A. W. Melton (Ed.), *Categories of human learning* (pp. 243-285). New York: Academic Press.
- FORT, A., DELPUECH, C., PERNIER, J., & GIARD, M.-H. (2002). Dynamics of cortico-subcortical cross-modal operations involved in audio-visual object detection in humans. *Cerebral Cortex*, **12**, 1031-1039.
- FOX, J. J., & SCHROEDER, C. E. (2005). The case for feedforward multisensory convergence during early cortical processing. *NeuroReport*, **16**, 419-423.
- FRENS, M. A., VAN OPSTAL, A. J., & VAN DER WILLIGEN, R. F. (1995). Spatial and temporal factors determine auditory-visual interactions in human saccadic eye movements. *Perception & Psychophysics*, **57**, 802-816.
- GESCHWIND, N. (1965). Disconnexion syndromes in animals and man: I. *Brain*, **88**, 237-294.
- GIARD, M.-H., & PERONNET, F. (1999). Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study. *Journal of Cognitive Neuroscience*, **11**, 473-490.
- GIELEN, S. C. A. M., SCHMIDT, R. A., & VAN DEN HEUVEL, P. J. M. (1983). On the nature of intersensory facilitation of reaction time. *Perception & Psychophysics*, **34**, 161-168.
- GIRAY, M., & ULRICH, R. (1993). Motor coactivation revealed by response force in divided and focused attention. *Journal of Experimental Psychology: Human Perception & Performance*, **6**, 1278-1291.
- GREEN, D. M., & SWETS, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- GREEN, K. P., STEVENS, E. B., KUHL, P. K., & MELTZOFF, A. M. (1990). Exploring the basis of the "McGurk effect": Can perceivers combine information from a female face and a male voice? *Journal of the Acoustical Society of America*, **87**, S125.
- GRICE, G. R., CANHAM, L., & BOROUGHS, J. M. (1984). Combination rule for redundant information in reaction time tasks with divided attention. *Perception & Psychophysics*, **35**, 451-463.
- HEIM, S., OPITZ, B., MÜLLER, K., & FRIEDERICI, A. D. (2003). Phonological processing during language production: fMRI evidence for a shared production-comprehension network. *Cognitive Brain Research*, **16**, 285-296.
- HOLMES, N. P., & SPENCE, C. (2005). Multisensory integration: Space, time and superadditivity. *Current Biology*, **15**, R762-R764.
- HUGHES, H. C., NELSON, M. D., & ARONCHICK, D. M. (1998). Spatial characteristics of visual-auditory summation in human saccades. *Vision Research*, **38**, 3955-3963.
- HUGHES, H. C., NOZAWA, G., & KITLERLE, F. (1996). Global precedence, spatial frequency, channels, and the statistics of natural images. *Journal of Cognitive Neuroscience*, **8**, 197-230.
- HUGHES, H. C., REUTER-LORENZ, P. A., NOZAWA, G., & FENDRICH, R. (1994). Visual-auditory interactions in sensorimotor processing: Saccades versus manual responses. *Journal of Experimental Psychology: Human Perception & Performance*, **20**, 131-153.
- HUGHES, H. C., & TOWNSEND, J. T. (1998). Varieties of binocular interaction in human vision. *Psychological Science*, **9**, 53-60.
- IACOBONI, M., WOODS, R. P., & MAZZIOTTA, J. C. (1998). Bimodal (auditory and visual) left frontoparietal circuitry for sensorimotor integration and sensorimotor learning. *Brain*, **121**, 2135-2143.
- ISA, T., & SASAKI, S. (2002). Brainstem control of head movements during orienting: Organization of the premotor circuits. *Progress in Neurobiology*, **66**, 205-241.
- JAY, M. F., & SPARKS, D. L. (1984). Auditory receptive fields in primate superior colliculus shift with changes in eye position. *Nature*, **309**, 345-347.
- JOASSIN, F., MAURAGE, P., BRUYER, R., CROMMELINCK, M., & CAMPANELLA, S. (2004). When audition alters vision: An event-related potential study of the cross-modal interactions between faces and voices. *Neuroscience Letters*, **369**, 132-137.
- KAYSER, C., PETKOV, C. I., AUGATH, M., & LOGOTHETIS, N. K. (2005). Integration of touch and sound in auditory cortex. *Neuron*, **48**, 373-384.
- KIM, Y. Y., LEE, B., SHIN, Y. W., KWON, J. S., & KIM, M. S. (2006). Ac-

- tivity of left inferior frontal gyrus related to word repetition effects: LORETA imaging with 128-channel EEG and individual MRI. *Neuro-Image*, **29**, 712-720.
- KIMURA, A., & TAMAI, Y. (1992). Sensory response of cortical neurons in the anterior ectosylvian sulcus, including the area evoking eye movement. *Brain Research*, **575**, 181-186.
- KRUMMENACHER, J., MÜLLER, H. J., & HELLER, D. (2001). Visual search for dimensionally redundant pop-out targets: Evidence for parallel-coactive processing of dimensions. *Perception & Psychophysics*, **63**, 901-917.
- LAURIENTI, P. J., BURDETTE, J. H., WALLACE, M. T., YEN, Y.-F., FIELD, A. S., & STEIN, B. E. (2002). Deactivation of sensory-specific cortex by cross-modal stimuli. *Journal of Cognitive Neuroscience*, **14**, 420-429.
- LAURIENTI, P. J., PERRAULT, T. J., STANFORD, T. R., WALLACE, M. T., & STEIN, B. E. (2005). On the use of superadditivity as a metric for characterizing multisensory integration in functional neuroimaging studies. *Experimental Brain Research*, **166**, 289-297.
- MACALUSO, E., GEORGE, N., DOLAN, R., SPENCE, C., & DRIVER, J. (2004). Spatial and temporal factors during processing of audiovisual speech: A PET study. *NeuroImage*, **21**, 725-732.
- MARINKOVIC, K., DHOND, R. P., DALE, A. M., GLESSNER, M., CARR, V., & HALGREN, E. (2003). Spatiotemporal dynamics of modality-specific and supramodal word processing. *Neuron*, **38**, 487-497.
- MAY, P. J. (2005). The mammalian superior colliculus: Laminar structure and connections. *Progress in Brain Research*, **151**, 321-378.
- MCGURK, H., & MACDONALD, J. (1976). Hearing lips and seeing voices. *Nature*, **264**, 746-748.
- MEREDITH, M. A., & STEIN, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, **221**, 389-391.
- MEREDITH, M. A., & STEIN, B. E. (1986a). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, **365**, 350-354.
- MEREDITH, M. A., & STEIN, B. E. (1986b). Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *Journal of Neurophysiology*, **56**, 640-662.
- MEREDITH, M. A., & STEIN, B. E. (1996). Spatial determinants of multisensory integration in cat superior colliculus. *Journal of Neurophysiology*, **75**, 1843-1857.
- MESULAM, M.-M. (1998). From sensation to cognition. *Brain*, **121**, 1013-1052.
- MEYER, G. F., WUERGER, S. M., RÖHRBEIN, F., & ZETTSCHKE, C. (2005). Low-level integration of auditory and visual motion signals requires spatial co-localisation. *Experimental Brain Research*, **166**, 538-547.
- MILLER, J. (1982). Divided attention: Evidence for coactivation with redundant signals. *Cognitive Psychology*, **14**, 247-279.
- MILLER, J. (1986). Timecourse of coactivation in bimodal divided attention. *Perception & Psychophysics*, **40**, 331-343.
- MILLER, L. M., & D'ESPOSITO, M. (2005). Perceptual fusion and stimulus coincidence in the cross-modal integration of speech. *Journal of Neuroscience*, **25**, 5884-5893.
- MORDKOFF, J. T., & YANTIS, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception & Performance*, **17**, 520-538.
- MOWBRAY, G. H. (1960). Choice reaction times for skilled responses. *Quarterly Journal of Experimental Psychology*, **7**, 193-202.
- MOWBRAY, G. H., & RHOADES, M. V. (1959). On the reduction of choice reaction times with practice. *Quarterly Journal of Experimental Psychology*, **6**, 16-23.
- MUNHALL, K. G., GRIBBLE, P., SACCO, L., & WARD, M. (1996). Temporal constraints on the McGurk effect. *Perception & Psychophysics*, **58**, 351-362.
- NAGY, M. E., & RUGG, M. D. (1989). Modulation of event-related potentials by word repetition: The effects of inter-item lag. *Psychophysiology*, **26**, 431-436.
- NEIL, P. A., CHEE-ROUTER, C., SCHEIER, C., LEWKOWICZ, D. J., & SHIMOJO, S. (2006). Development of multisensory spatial integration and perception in humans. *Developmental Science*, **9**, 454-464.
- NICKERSON, R. S. (1973). Intersensory facilitation of reaction time: Energy summation or preparation enhancement? *Psychological Review*, **80**, 489-509.
- NOZAWA, G., REUTER-LORENZ, P., & HUGHES, H. C. (1994). Parallel and serial processes in the human oculomotor system: Bimodal integration and express saccades. *Biological Cybernetics*, **72**, 19-34.
- OLSON, I. R., GATENBY, J. C., & GORE, J. C. (2002). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. *Cognitive Brain Research*, **14**, 129-138.
- PATCHING, G. R., & QUINLAN, P. T. (2004). Cross-modal integration of simple auditory and visual events. *Perception & Psychophysics*, **66**, 131-140.
- PATTON, P., BELKACEM-BOUSSAID, K., & ANASTASIO, T. J. (2002). Multimodality in the superior colliculus: An information theoretic analysis. *Cognitive Brain Research*, **14**, 10-19.
- POPULIN, L. C., & YIN, T. C. (2002). Bimodal interactions in the superior colliculus of the behaving cat. *Journal of Neuroscience*, **22**, 2826-2834.
- RAAB, D. H. (1962). Statistical facilitation of simple reaction times. *Transactions of the New York Academy of Sciences*, **24**, 574-590.
- RAJ, T., UUTELA, K., & HARI, R. (2000). Audiovisual integration of letters in the human brain. *Neuron*, **28**, 617-625.
- REGAN, J. E. (1981). Automaticity and learning: Effects of familiarity on naming letters. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 180-195.
- SAITO, D. N., YOSHIMURA, K., KOCHIYAMA, T., OKADA, T., HONDA, M., & SADATO, N. (2005). Cross-modal binding and activated attentional networks during audio-visual speech integration: A functional MRI study. *Cerebral Cortex*, **15**, 1750-1760.
- SANTI, A., SERVOS, P., VATIKIOTIS-BATESON, E., KURATATE, T., & MUNHALL, K. (2003). Perceiving biological motion: Dissociating visible speech from walking. *Journal of Cognitive Neuroscience*, **15**, 800-809.
- SCHACTER, D. L., & BUCKNER, R. L. (1998). Priming and the brain. *Neuron*, **20**, 185-195.
- SCHNEIDER, K. A., & KASTNER, S. (2005). Visual responses of the human superior colliculus: A high-resolution functional magnetic resonance imaging study. *Journal of Neurophysiology*, **94**, 2491-2503.
- SPARKS, D. L. (1986). Translation of sensory signals into commands for control of saccadic eye movements: Role of primate superior colliculus. *Physiology Review*, **66**, 118-171.
- STEIN, B. E. (1978). Development and organization of multimodal representation in cat superior colliculus. *Federation Proceedings*, **37**, 2240-2245.
- STEIN, B. E. (1998). Neural mechanisms for synthesizing sensory information and producing adaptive behaviors. *Experimental Brain Research*, **123**, 124-135.
- STEIN, B. E., & ARIGBEDE, M. O. (1972). Unimodal and multimodal response properties of neurons in the cat's superior colliculus. *Experimental Neurology*, **36**, 179-196.
- STEIN, B. E., JIANG, W., WALLACE, M. T., & STANFORD, T. R. (2001). Nonvisual influences on visual-information processing in the superior colliculus. *Progress in Brain Research*, **134**, 143-156.
- STEIN, B. E., & MEREDITH, A. M. (1993). *The merging of the senses*. Cambridge, MA: MIT Press.
- STEIN, B. E., WALLACE, M. T., STANFORD, T. R., & JIANG, W. (2002). Cortex governs multisensory integration in the midbrain. *Neuroscientist*, **8**, 306-314.
- SUMBY, W. H., & POLLACK, I. (1954). Visual contributions to speech intelligibility in noise. *Journal of the Acoustical Society of America*, **26**, 212-215.
- TALSMA, D., & WOLDORFF, M. G. (2005). Selective attention and multisensory integration: Multiple phases of effects on the evoked brain activity. *Journal of Cognitive Neuroscience*, **17**, 1098-1114.
- THEIOS, J. (1975). The components of response latency in simple human information processing tasks. In P. M. A. Rabbitt & S. Dornic (Eds.), *Attention and performance V* (pp. 418-440). New York: Academic Press.
- TODD, J. W. (1912). *Reaction to multiple stimuli* (Archives of Psychology, Vol. 25). New York: Science Press.
- TOLDI, J., FEHÉR, O., & WOLFF, J. R. (1986). Sensory interactive zones in the rat cerebral cortex. *Neuroscience*, **18**, 461-465.
- TOWNSEND, J. T., & ASHBY, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.
- TOWNSEND, J. T., & NOZAWA, G. (1995). On the spatio-temporal properties of elementary perception: An investigation of parallel, serial and coactive theories. *Journal of Mathematical Psychology*, **39**, 321-360.
- TOWNSEND, J. T., & WENGER, M. J. (2004). A theory of interactive paral-

- lel processing: New capacity measures and predictions for a response time inequality series. *Psychological Review*, **111**, 1003-1035.
- TURKELTAUB, P. E., EDEN, G. F., JONES, K. M., & ZEFFIRO, T. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, **16**, 765-780.
- VAN ATTEVELDT, N. M., FORMISANO, E., BLOMERT, L., & GOEBEL, R. (2007). The effect of temporal asynchrony on the multisensory integration of letters and speech sounds. *Cerebral Cortex*, **17**, 962-974.
- WALLACE, M. T., MEREDITH, M. A., & STEIN, B. E. (1992). Integration of multiple sensory modalities in cat cortex. *Experimental Brain Research*, **91**, 484-488.
- WALLACE, M. T., & STEIN, B. E. (2001). Sensory and multisensory responses in the newborn monkey superior colliculus. *Journal of Neuroscience*, **21**, 8886-8894.
- WOODWORTH, R. S., & SCHLOSBERG, H. (1954). *Experimental psychology* (Rev. ed.). New York: Holt, Rinehart & Winston.
- WUERGER, S. M., HOFBAUER, M., & MEYER, G. F. (2003). The integration of auditory and visual motion signals at threshold. *Perception & Psychophysics*, **65**, 1188-1196.

(Manuscript received November 21, 2005;
revision accepted for publication October 9, 2006.)