INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6" x 9" black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

ProQuest Information and Learning
300 North Zeeb Road, Ann Arbor, MI 48106-1346 USA
800-521-0600

UMI®

Reproduced with permission of the copyright owner. Further reproduction prohibited without permission.

THE CORTICAL NEUROPHYSIOLOGY OF VISUAL-AUDITORY MULTISENSORY PROCESSING IN HUMANS

By
Sophie Molholm

A dissertation submitted to the Graduate Faculty in Psychology in partial fulfillment of
the requirements for the degree of Doctor of Philosophy,
The City University of New York

2002

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Psychology in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

4-29-02
Date

Walter Ritter
Chair of Examining Committee

5/1/02
Date

Executive Officer

Walter Ritter

John J. Foxe

Diana Deacon

Charles Spence

David Friedman

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

THE CORTICAL NEUROPHYSIOLOGY OF VISUAL-AUDITORY MULTISENSORY PROCESSING IN HUMANS

by

Sophie Molholm

Advisor: Walter Ritter

This research investigated the spatio-temporal dynamics of the cortical integration of inputs to the visual and auditory sensory systems, with the use of high-density electrical scalp recordings. The first study examined visual-auditory multisensory neural interactions for basic visual and auditory stimuli under the most simple of tasks. The spatio-temporal dynamics of multisensory neural interactions were mapped over the first 200 ms of cortical processing post-stimulus onset, and compared to ongoing unisensory processing. The second study investigated the spatio-temporal dynamics of multisensory influences on object recognition processes, by comparing multisensory processing of visual and auditory elements that belonged to the same common object or belonged to different common objects (i.e., line drawings or animals and animal vocalizations from the same or different animals), for stimuli that were task relevant versus stimuli that were task irrelevant. Overall, the data from these studies indicate that multisensory interactions occur at many levels of processing, and can be affected by task parameters and the semantic content of the different sensory elements of multisensory stimuli. Further, the data reveal that there is considerable multisensory processing in sensory specific cortical areas, and suggest that in some cases this occurs in a feed-forward manner.

Dedication

This is dedicated to my husband Adil and my son Aidan, whom I love very much.

Acknowledgements

I am grateful to have had the good fortune to work with excellent scientists who take their role in mentoring seriously and do it exceptionally well. In particular both Walter Ritter and John Foxe have contributed considerably to my development as a scientist. It is not possible to account for all that I have gained from knowing Walter, but suffice it to say that he goes above and beyond the call of duty. Working with him is always interesting and often a lot of fun as well. John has contributed significantly to my approach to cognitive neuroscience, and I thank him for his mentorship and providing me the opportunity to achieve something that I am proud of.

Many thanks to John Antrobus, it was in his Cognition course, pre-graduate student career, that I first realized a fascination with the perceptual/cognitive sciences. Thanks also to Vicky Tartter; her engaging and informative graduate courses were inspiring and furthered this interest. Thanks to Adam Anderson and Eve DeRosa for being dear people, for sharing their graduate school experiences with me, and for encouraging me all along. Thanks to Hilary Gomes for her support and friendship, and for helping me to keep things in perspective.

Thanks to everyone in the Cognitive Neurophysiology Laboratory and Life Sciences department at the Nathan Kline Institute, who make it a stimulating and supportive environment to do research in. This includes Dan Javitt, the director of the Life Sciences department; John Foxe, the director of the Cognitive Neurophysiology Laboratory; Micah Murray (now at the University Hospital of Geneva), a friendly face

who showed me the ropes when I first arrived at the laboratory; Glenn Wylie, who frequently dropped by my office to check up on me and put a smile on my face these last couple anxious months; Deirdre Foxe and Beth Higgins, who are excellent people as well as excellent technicians; and Charlie Schroeder, Antigona Martinez and Elisa Dias, who add significantly to this exceptional mix of people. Last, but certainly not least, thanks to my husband Adil, my mother Nina, and the rest of my wonderful family, for I couldn't have done this without them.

Table of Contents

CHAPTER 1

Introduction

Inputs to the various sensory systems function to inform us of the presence and identity of the objects and events in the environment at any given moment, and our experience is one in which the multiple sensory properties cohere into unified perceptions of these object and events. To illustrate this, it helps to imagine what it is like when we are deprived of the contribution of one of these senses. For example, when it is dark we have to rely on auditory and somatosensory information in the near absence of visual information. In this case our experience is dramatically different, and objects and events can seem incomplete. The same would be the case were we deprived of auditory or somatosensory stimulation. Thus while we easily distinguish the unique characteristics of different sensory inputs (e.g., the touch, smell, and image of a flower), phenomenological experience, as well as experimental evidence (e.g., Stein, Huneycutt, & Meredith, 1988; Bertelson & Radeau, 1981; McGurk & MacDonald, 1976; MacGrath & Summerfield, 1985), tells us that there are profound multisensory influences on the normal perception of objects and events.

Behavioral evidence of multisensory perception

A number of behavioral studies have demonstrated and to some extent delineated multisensory influences on the detection, localization, and identification of stimuli. These have shown that the presence of multisensory cues can affect both the detection and localization of a stimulus (e.g., Stein, Huneycutt, & Meredith, 1988; Bertelson & Radeau, 1981), and that multisensory cues can combine to improve object recognition or to affect the perceived identity of an object (e.g., MacGrath & Summerfield, 1985; McGurk & MacDonald, 1976).

The perceived intensity of a stimulus can be increased by the simultaneous presentation of a stimulus in another sensory modality (Stein, London, Wilkinson, & Price, 1996), such that the detection of a faint stimulus is improved (Stein et al., 1988). Stein and colleagues (Stein et al., 1988) presented brief flashes and tones alone or simultaneously to cats that were trained to detect a visual stimulus, and ignore the auditory stimuli. The flashes and tones were presented at matching or mismatching locations of different eccentricities relative to the cat. For visual stimuli that were detected less than 100% of the time when presented alone, detection was improved when an auditory stimulus was simultaneously presented to the same location.

The presence of multiple sensory cues has been shown to both enhance and interfere with the accurate localization of a stimulus. Spatial acuity improves with location concordant bisensory stimulation, as demonstrated by the effects of concurrent auditory or somatosensory stimulation on the localization of visual stimuli (see Welch & Warren 1986 for a review; Stein & Meredith, 1993). On the other hand, spatial acuity can decrease with location discordant bisensory stimulation (e.g., Bertelson & Radeau, 1981). This latter phenomenon is referred to as 'the ventriloquist effect', so called because it is best illustrated by conjuring to mind the talking dummy with his 'master' the ventriloquist standing by.

Multisensory effects on object/event perception have been largely demonstrated for speech perception, showing that viewing articulatory gestures can affect what speech sound is perceived (e.g., McGurk & McDonald, 1976), or enhance speech perception in a noisy environment (Campbell & Dodd, 1980; Thompson, 1995). The McGurk effect (McGurk & MacDonald, 1976) exemplifies the multisensory nature of speech perception.

In one demonstration, the speech sound “ba” was presented and unambiguously heard as “ba”. When it was paired with a video of someone mouthing the syllable “ga”, “da” or “tha” were most often heard. These speech syllables fall between “ba” and “ga” in terms of place of articulation (Delattre, Liberman, & Cooper, 1955), suggesting that visual and auditory speech information were fused to arrive at a final speech percept. This highly robust illusion provides a clear example of the multisensory nature of speech perception.

Related to multisensory object perception, there are several studies that have looked at multisensory influences on the perception of features. In a clever experiment, Jousmäki and Hari (1998) showed that the perception of surface texture was influenced by simultaneous auditory stimulation, and in this way demonstrated that both tactile and auditory information can contribute to the perception of texture. Subjects rubbed the palms of their hands together, and the sound their hands made was played back to them online, over headphones. When the higher frequencies were amplified, subjects rated their hands as feeling drier and smoother than when they were not modified. In another study, visual and tactile cues were shown to both contribute to the perception of height (Ernst & Banks, 2002).

Thus we see that inputs to the multiple sensory channels combine to affect perception at many levels, i.e., the detection and localization of stimuli, and the perception of stimulus features and of objects and events.

Neural mechanisms of multisensory integration

Most of what we know about the mechanisms of neural multisensory integration comes from the work of Stein and colleagues (see Stein & Meredith, 1993 for a review) on multisensory integration in cells of the superior colliculus. This has been largely investigated in cats for combinations of visual, auditory, and somatosensory stimulation, using the method of single cell recordings. Single cell multisensory integration is defined as a change in cell responsiveness to stimulation by two or more sensory inputs, compared to the sum of the responses to the different inputs under conditions of unisensory stimulation. This can be either an increase or a decrease in response rate. Based on the work of Stein and colleagues several principles of multisensory integration have emerged, for cells in the superior colliculus: 1) stimuli must be in temporal synchrony (within ~100 ms; Meredith, Nemitz, & Stein, 1987); 2) stimuli must be in spatial register, such that they fall into the receptive field of the same multisensory cell (Meredith & Stein, 1996); 3) multisensory interactions are strongest when the individual inputs are minimally effective (the principle of inverse effectiveness; Stein & Meredith, 1993); and 4) stimuli that are out of spatial register can result in cell response depression (Meredith & Stein, 1996).

Recent evidence suggests that cortical neurons in known polysensory areas follow the principles of multisensory integration defined for the superior colliculus, with the exception that spatial register seems not to be as important (e.g., Duhamel et al., 1991; Wallace et al., 1992). Data from our laboratory supports the notion that spatial register is not an essential condition for eliciting cortical multisensory interactions. Using the method of high-density event-related potentials (ERP), we found that multisensory interactions were elicited by auditory and somatosensory inputs that were out of spatial register by 180 degrees.

Electrophysiological and imaging data on multisensory processing

Knowledge of the cortical spatio-temporal network of multisensory processing is an essential component to the full understanding of multisensory influences on perception. Event-related potentials (ERP) provide precise temporal information on the time course of cortical processes, as well as and some insight into the possible underlying generators of cortical processes; while functional imaging provides precise detail of the cortical structures that are activated by a given process. There are a number of functional imaging studies on multisensory processing, and a few such electrophysiological studies. Evidence from electrophysiological and functional data combine to strongly suggest that multisensory processing occurs early on in information processing, in cortical areas that are generally thought to be 'unisensory' (Calvert et al., 1997; Foxe et al., 2000; Giard & Peronnet, 1999; Macaluso et al., 2000). Such early multisensory effects, which appear to occur in cortical areas involved in the initial sensory processing of the individual sensory inputs, may be related to the detection and localization of stimuli, as well as possibly 'tagging' inputs to the different sensory systems as belonging to the same object. The functional and electrophysiological data also suggest that cortical multisensory processing is affected by the semantic or emotional congruency of the sensory inputs (Calvert et al., 1999; 2000; de Gelder et al., 1999; Pourtois et al., 2000; Raij, Uutela, & Hari, 2000), in both polysensory and 'unisensory' areas. The little data that are currently available on the timing of such effects vary considerably, probably in part due to the very different stimuli and tasks used across studies (de Gelder et al., 1999; Pourtois et al., 2000; Raij, Uutela, & Hari, 2000). The lower estimates (~100 ms) come from studies that suffer from methodological problems (de Gelder et al., 1999; Pourtois et al., 2000: The problem is that the visual stimulus was presented well before the auditory stimulus (~1000 ms) and stayed on through the duration of the auditory stimulus, thus effects could be due to anticipatory processes), while the estimate on the higher end (280 ms) may have been due to the use of letters (Raij, Uutela, & Hari, 2000), which presumably require more translation to become meaningful than, say, clear images of animals.

Thus a picture of cortical multisensory processing is beginning to emerge, which suggests that inputs to separate sensory channels that are temporally coincident result in the early multisensory processing in cortical areas that are involved in the sensory processing of the stimuli.

The present endeavor

The goal of the research presented in this thesis was to investigate the spatio-temporal dynamics of the cortical integration of inputs to the visual and auditory sensory systems, with the use of high-density electrical scalp recordings. The first study (Chapter 2) examined visual-auditory multisensory neural interactions for basic visual and auditory stimuli under the most simple of tasks. The spatio-temporal dynamics of multisensory neural interactions were mapped over the first 200 ms of cortical processing post-stimulus onset. Particular consideration was given to the point in time at which such interactions first onset, and the possible underlying neural generators of this initial multisensory interaction, given its scalp topography. This study provided a full description of the time course of multisensory processing for visual-auditory stimuli that originated from the same location in space, onset at the same time, and were task relevant (a simple reaction-time task was employed). However, beyond these factors the stimuli were not related in any ecologically meaningful way.

The second study (Chapter 2) was designed to investigate the spatio-temporal dynamics of multisensory object recognition processes. This was achieved by comparing multisensory processing of visual and auditory elements that belonged to the same object or belonged to different objects (i.e., line drawings or animals and animal vocalizations from the same or different animals), for stimuli that were task relevant versus stimuli that were task irrelevant. Thus the study examined the points in information processing, and the general cortical areas, that stimuli with congruent versus incongruent visual and auditory elements were differentiated, with the purpose of placing these effects in the framework of ongoing information processing.

CHAPTER 2

Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study.

Reprinted from Cognitive Brain Research, 14, Molholm, S., Ritter, W., Murray, M., Javitt, D.C., Schroeder, C.E., & Foxe, J.J., Multisensory visual auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study, Copyright (2002), with permission from Elsevier Science.

Abstract: Integration of information from multiple senses is fundamental to perception and cognition, but when and where this is accomplished in the brain is not well understood. This study examined the timing and topography of cortical auditory-visual interactions using high-density event-related potentials (ERPs) during a simple reaction-time (RT) task. Visual and auditory stimuli were presented alone and simultaneously. ERPs elicited by the auditory and visual stimuli when presented alone were summed ('sum' ERP) and compared to the ERP elicited when they were presented simultaneously ('simultaneous' ERP). Divergence between the 'simultaneous' and 'sum' ERP indicated auditory-visual (AV) neural response interactions. There was a surprisingly early right parieto-occipital AV interaction, consistent with the finding of an earlier study (M.H. Giard, F. Peronnet, Journal of Cognitive Neuroscience 11 (1999) 473-490). The timing of onset of this effect (46 ms) was essentially simultaneous with the onset of visual cortical processing, as indexed by the onset of the visual C1 component, which is thought to represent the earliest cortical visual evoked potential. The coincident timing of the early AV interaction and C1 strongly suggests that AV interactions can affect early visual sensory processing. Additional AV interactions were found within the time course of sensory processing (up to 200 ms post stimulus onset). In total, this system of AV effects over the scalp was suggestive of both activity unique to multisensory processing, and the modulation of "unisensory" activity. RTs to the stimuli when presented simultaneously were significantly faster than when they were presented alone. This RT facilitation could not be accounted for by probability summation, as evidenced by violation of the 'race' model, providing compelling evidence that auditory-visual neural interactions give rise to this RT effect.

1. Introduction

Everyday tasks involve the seemingly automatic integration of information from multiple sensory modalities. For instance, driving a car involves the synthesis of visual (seeing the road), auditory (hearing the car engine; a passing car), somatosensory (feeling the steering wheel), and motor (depressing the gas pedal) activity. The combination of inputs from different senses can function to reduce perceptual ambiguity (e.g., 32) and enhance stimulus detection (e.g., 53). Despite the fundamental role that sensory integration plays in performance and perception, how and when information from separate sensory modalities comes together in the human neocortex is not well understood. The bulk of our knowledge on the mechanisms of multisensory integration in the brain comes from the pioneering research of Stein, Meredith, and colleagues (see 54 for a review) on multisensory processing in the superior colliculus (of anaesthetized cats), a sub-cortical structure involved in orienting to auditory, visual, and somatosensory stimuli. However, the extent to which the multisensory mechanisms defined in the superior colliculus generalize to cortical processes remains to be fully elucidated. Knowledge of the timing and anatomical distribution of cortical multisensory processing is essential to determining the roles that it plays in information processing.

Generally, it has been assumed that cortical multisensory processing occurs relatively late, following extensive processing of sensory inputs, and that it occurs in higher order cortical areas specialized for this purpose (e.g. 33). This assumption can be partially attributed to: 1) a bias resulting from the tradition of studying sensory systems in isolation, and 2) animal studies that reveal multisensory convergence in higher-order regions of the parietal (e.g. 17, 28, 50), temporal (e.g. 5, 12, 27), and frontal lobes (e.g. 3,

61) along with a general lack of corresponding studies demonstrating convergence in lower-tier cortical areas. However, recent evidence suggests that multisensory processing occurs during initial sensory transmission, and in cortical areas that are usually held to be unisensory. An investigation by Schroeder et al (47) in the caudomedial (CM) belt area of the auditory association cortex of awake behaving macaque monkeys, which gets direct input from primary auditory cortex (A1), showed auditory-somatosensory co-representation. Critically, both the auditory and somatosensory inputs to CM had characteristic feed-forward patterns, with both inputs arriving first in layer 4 at about 12 ms post stimulus onset, strongly suggesting bottom-up multisensory integration that occurs early in the sensory processing hierarchy. Functional imaging studies have suggested multisensory effects in what have been classically considered unisensory cortical areas (7, 31), although the prevailing opinion is that these interactions represent feedback from higher-tier multisensory onto the lower-tier unisensory areas. Direct empirical evidence of feedback mediated multisensory convergence in classical sensory cortex is sparse but supports this possibility (see 49). Very recently two event-related potential (ERP) studies found surprisingly early multisensory effects that, in light of their scalp topography, appear to indicate the early integration of sensory information in traditionally held unisensory cortex. In Giard and Peronnet (24), auditory-visual (AV) effects were found to onset at just 40 ms over right parieto-occipital scalp; this is consistent with generators in early visual cortices, although the spatial resolution of ERPs does not allow the contribution of the abutting multisensory areas in posterior parietal cortex or superior temporal sulcus (STS), to be ruled out. And in Foxe et al. (20), auditory-somatosensory effects onset at about 50 ms over central/post-central scalp,

consistent with generators in early somatosensory cortex; and at 70 ms over scalp areas consistent with neural activity from posterior auditory areas, in line with the findings of Schroeder et al. (47).

The finding of an AV effect that onsets at 40 ms over parieto-occipital scalp (24) suggests that AV effects can occur at about the same time that initial activation of primary visual cortex (V1) is usually assumed to occur (45-55 ms as represented by the onset of the earliest cortical visual evoked potential, C1: e.g. 10, 11, 21). This surprisingly early latency finding suggests a model of auditory-visual interaction in the cortex where auditory input, which reaches the cortex in less than half the time of visual input (9 to 15 ms: 9, 59), is transmitted from auditory cortices to visual or nearby visually dominant cortical areas, and consequently affects the early sensory processing of visual input.

The purpose of the present study was to advance the understanding of cortical multisensory processing by placing early AV interactions within the temporal and topographical framework of cortical sensory processing of the individual auditory and visual inputs. We first endeavored to determine if the early AV effect reported in Giard and Peronnet (24) would be elicited using a simple task and basic stimuli. In Giard and Peronnet a relatively complicated task and stimulus set were employed: On each trial, one of two tones was presented, and/or a permanently placed circle morphed into a horizontal or vertical ellipse. Subjects made forced two-choice classifications of the six randomly occurring stimulus conditions. In contrast, in the present study, single visually presented disks and auditory pure tones were presented either alone or simultaneously, and subjects performed a speeded simple reaction-time task. Elicitation of the effect under these conditions, in conjunction with its elicitation in the very different conditions of Giard and Peronnet (24), would suggest that early cortical multisensory processing for auditory and visual stimuli that onset simultaneously may be present for a variety of stimuli and tasks. We then compared the onset of the earliest AV effect to the onset of C1 in response to the visual stimulus alone. We expected that the initial cortical response to the visual stimulus would precede any AV effects, reflecting that cortical unisensory processing of the visual input began prior to cortical multisensory processing.

The technique of high-density electrical mapping (from 128 scalp electrodes) was used to establish the spatio-temporal dynamics of auditory-visual multisensory processing in relation to activation across a distributed sensory processing network. To assess multisensory processing, ERPs to the 'visual alone' and 'auditory alone' stimulus conditions were summed (hereafter referred to as the 'sum' ERP) and compared to the ERP to the simultaneously presented auditory and visual stimuli (the 'simultaneous' ERP). If neural responses to the auditory and visual inputs were processed in the same way when they were presented simultaneously as when they were presented alone, then, based on the principle of superposition of electrical fields, the 'simultaneous' ERP would be equivalent to the 'sum' ERP. However, if the neural responses to the simultaneously presented auditory and visual stimuli interacted during processing, the 'simultaneous' and 'sum' ERPs would diverge. This method of measuring multisensory processing is valid when neural responses reflect sensory processing unique to the stimulus, and do not reflect processes common to all three stimulus types such as target (e.g., the P3) or response (e.g., motor cortex activity) related neural activity. Several forms of interaction effects have been reported from this comparison (e.g. 20, 24, 38, 40, 55). Although our primary focus was on the earliest AV interaction, AV interactions up to 200 ms were considered.

We also tested whether multisensory processing was reflected in our behavioral data. Simple reaction-times are generally facilitated when location concordant stimuli are simultaneously presented. This has been called the "redundant signal effect" (RSE) (e.g., 29, 37). There are two classes of models to explain this effect: race models and coactivation models. In race models each stimulus of a pair *independently* competes for response initiation, and the faster of the two mediates the response for any trial. According to this model probability summation produces the RSE, since the likelihood of either of the two stimuli yielding a fast reaction-time is higher than that from one stimulus alone. In coactivation models, the *interaction* of neural responses to the simultaneously presented stimuli facilitates response initiation and produces the RSE. We tested whether the RSE exceeded the statistical facilitation predicted by the race model, and thereby provided evidence for the contribution of AV neural interactions to RT facilitation.

2. Materials and Methods

2.1. Subjects

Twelve neurologically normal, paid volunteers participated (mean age 23.8 ± 2.69 SD; 5 female; 11 right-handed), all reported normal hearing and normal or corrected-to-normal vision. The Institutional Review Board of the Nathan Kline Institute for Psychiatric Research approved the experimental procedures, and each subject provided written informed consent. Data from two additional subjects were excluded, one for excessive blinking, and the other for failure to maintain central fixation.

2.2. Stimuli

Auditory alone: a 1000 Hz tone (60 ms duration; 75 dB SPL) was presented from a single JBL speaker located atop the monitor on which visual stimulus was presented.

Visual alone: a disk (60 ms duration), subtending 1.2° in diameter (143 cm viewing distance) and appearing red on a black background, was presented on a monitor (Iiyama Vision Master Pro 502, model #A102GT), peripheral to central fixation. The location of the circle was chosen on an individual subject basis after completion of the following procedure. In a pre-test session, the red circle was presented 100 times to each of 16 different locations across the visual field. The center of the stimulus was located 1.6 or 2.6 degrees lateral from fixation and 1.6 or 2.0 degrees vertical from fixation. For each subject, the location was chosen from which the best C1-P1 visual evoked potentials (VEPs) were observed. We sought to optimize our sensitivity to the obligatory VEPs because cortical AV effects have been previously reported in the latency range and general scalp location of C1 and P1 (40 to 90 ms post-stimulus; 24). There were four different locations used across our cohort (see Figure 1).

Auditory and Visual simultaneous: the auditory-alone and visual-alone conditions described above were presented simultaneously. The auditory and visual stimuli were in close proximity, with the speaker placed on top of the monitor in vertical alignment with the visual stimulus.

2.3. Procedure

Participants were seated in a comfortable chair in a dimly lit and electrically shielded (Braden Shielding Systems) room and asked to keep head and eye movements to a minimum, while maintaining central fixation. Eye position was monitored with horizontal and vertical electro-oculogram (EOG) recordings. Subjects were instructed to make a button-press response with their right index finger when a stimulus in either sensory modality was detected, as quickly as possible without making errors. The three stimulus conditions were presented with equal probability in random order. Stimulus onset asynchrony (SOA) varied randomly between 750 and 3000 ms. Stimuli were blocked into sequences of 150 trials, and each subject completed a minimum of 18 blocks. Breaks were encouraged between blocks to maintain high concentration and prevent fatigue.

2.4. Data acquisition and analysis

Continuous EEG was acquired from 128 scalp electrodes (impedances < 5kΩ), referenced to the nose, band-pass filtered from 0.05 to 100 Hz, and digitized at 500 Hz. The continuous EEG was divided into epochs (-100 ms pre- to 800 ms post- stimulus onset) and baseline corrected over the full 900 ms. Trials with blinks and eye movements were rejected off-line on the basis of the EOG. An artifact criterion of ± 60 µV was used at all other scalp sites to reject trials with excessive EMG or other noise transients. The average number of accepted sweeps per stimulus condition was 681 (ranging from 661 to 702 sweeps over the three conditions)

EEG epochs were sorted according to stimulus condition and averaged from each subject to compute the event-related potential (ERP). Baseline was defined as the epoch from –50 ms to 10 ms post-stimulus onset. In addition, ERPs from the auditory-alone and visual-alone conditions were summed for statistical comparison with the ERP response to the simultaneous auditory-visual condition. Group-averaged ERPs for each of the three stimulus conditions and the summed auditory and visual alone ERPs were calculated for display purposes and for identification of the auditory N1 and P2 and the visual C1, P1, and N1.

Button press responses to the three stimulus conditions were acquired during the recording of the EEG and processed offline. Responses falling between 100 and 800 ms post stimulus onset were considered valid. This window was used to avoid the double categorization of a response.

2.5. Statistical analyses

2.5.1. Behavioral. For individual subjects, the percent hits and average response-time were calculated and their reaction-time distributions were recorded for each stimulus condition. To test for the presence of an RSE, planned comparisons between each of the alone stimulus conditions and the simultaneous stimulus condition were performed. To determine if a co-activation explanation of the RSE was necessary, Miller's test of the race model (37) was implemented. The model places an upper limit on the cumulative probability (CP) of RT at a given latency for stimulus pairs. For any latency, t, the race model holds when this CP value is less than or equal to the sum of the CP from each of the single stimuli minus an expression of their joint probability

($CP_{(t)simultaneous} \leq ((CP_{(t)alone\ 1} + CP_{(t)alone\ 2}) - (CP_{(t)alone\ 1} * CP_{(t)alone\ 2}))$. For each subject the RT range within the valid RTs (100-800 ms) was calculated over the three stimulus conditions and divided into quantiles from the first to the hundredth percentile in 5% increments (1%, 5%, ..., 95%, 100%). T-tests comparing the actual ($CP_{(t)simultaneous}$) and predicted ($(CP_{(t)alone\ 1} + CP_{(t)alone\ 2}) - (CP_{(t)alone\ 1} * CP_{(t)alone\ 2})$) facilitation were performed to assess the reliability of violations of the race model.

2.5.2. Event-related potentials: AV effects. To date only Giard and Peronnet (24) have measured AV multisensory processing by comparing 'simultaneous' and 'sum' ERPs. In the first phase of our analysis we used their findings to predefine the latencies and scalp regions at which we tested for AV interactions. As such, we tested for an early AV interaction over parieto-occipital scalp within the latency range of the C1, AV interactions within the latency range of the auditory N1 and P2 and the visual P1 and N1 components, over corresponding scalp areas, and an AV effect over right fronto-temporal scalp at about 150 ms. For each of the six tests the mean amplitude of the 'simultaneous' and 'sum' ERPs over a 10 ms window was submitted to a three-way repeated measures analysis of variance (ANOVA) with factors of Stimulus Type ('simultaneous' versus 'sum'), Hemisphere, and Electrode (3 representing each hemisphere). The center of each of the latency windows was chosen on the basis of the maximum difference between the 'simultaneous' and 'sum' ERPs within the prescribed latency range, and is indicated in parenthesis in the results section. Geisser-Greenhouse corrections were used in reporting *p* values when appropriate. An alpha level of less than .05 was used for all statistical tests. For the reliable AV interactions, the topography and distribution of the effect was inspected in the difference voltage map; when the focus of the effect differed from the scalp region tested, a follow up test was conducted.

The second phase of analysis was conducted to render a full description of the spatio-temporal properties of the AV interactions, and was considered exploratory. The 'simultaneous' and 'sum' ERPs were compared using point-wise running paired t-tests (2-tailed), with an AV interaction defined as at least 10 consecutive data points meeting a 0.05 alpha criterion (10 data points = 20 ms at a 500 Hz digitization rate) (for use of running t-tests see 26, 40). This criterion meets stringent standards for assessing reliable effects, as posed by Guthrie and Buchwald (26), when a large number of t-tests are calculated across the electrode montage and epoch. This is a suitable alternative to Bonferroni correction for multiple comparisons, which would increase the likelihood of Type II errors through overcompensation for Type I errors. This "post-hoc" phase of the analysis was critical because 1) it provided a description of the time course of the AV effects, and 2) the full description of the spatio-temporal properties of the AV interactions allows these data to be used as a hypothesis generation tool for future multisensory studies.

2.5.3. Event-related potentials: C1 onset. C1 onset was measured by comparing the amplitude of the ERP to the visual alone condition to baseline using point-wise running paired t-tests (2-tailed); onset was defined as the first point where a 0.05 alpha criterion was met and followed by at least 10 consecutive significant data points.

3. Results

3.1. Behavioral

3.1.1. Redundant signal effect. Mean reaction-times to the simultaneous condition (255 ms) were faster than mean reaction-times to either the visual or the auditory alone conditions (305 and 297, respectively). A RSE was confirmed with planned comparisons of each of the alone conditions to the simultaneous condition (for auditory alone condition compared to the simultaneous condition: $t_{11} = 7.972$, $p < 0.001$; for the visual alone condition compared to the simultaneous condition: $t_{11} = 11.057$, $p < 0.001$).

3.1.2. Test of the race model. There was violation of the race model for each of the 12 subjects. The CP at each quantile were group averaged separately for each stimulus condition to form an aggregate distribution that preserved the shape of the individuals' data (also termed Vincent averaging; 60). For the group averaged data, the race model was violated over the first 6 quantiles (see Figure 2). These violations were reliable across subjects, with statistically significant differences between the actual ($CP_{(t)simultaneous}$) and predicted ($(CP_{(t)alone\ 1} + CP_{(t)alone\ 2}) - (CP_{(t)alone\ 1} * CP_{(t)alone\ 2})$) facilitation for three of the quantiles (see Table 1). Violation of the race model strongly supports the contention that the interaction of auditory and visual stimulus information during neural processing contributed to the RSE.

3.2. Event-related potentials

In Figure 3 the group averaged ERPs to the alone conditions (auditory and visual), and the simultaneous condition at selected electrode sites are displayed. The ERP elicited by the auditory alone condition was characterized by the typical auditory components Pa peaking at 40 ms, P1 peaking at 56 ms, N1 peaking at 105 ms, and P2 peaking at about 200 ms (43, 57, 58). Pa and N1 appeared maximal over fronto-central scalp and inverted at the mastoids; P1 appeared maximal at fronto-central sites but did not invert at the mastoids; and P2 appeared maximal over centro-parietal scalp. The ERP elicited by the visual alone condition was characterized by the visual C1 from about 40 to 65 ms, and a typical P1 and N1 peaking at 110 and 160 ms, respectively (10, 21); P1 appeared maximal over occipital scalp and N1 over occipito-temporal scalp. The ERP elicited by the 'simultaneous' condition displayed classic auditory and visual componentry, in general as described for the alone conditions.

3.2.1. AV Interactions

Differences between the group averaged 'simultaneous' and 'sum' ERPs could be readily observed. For the electrode sites over parieto-occipital to occipital scalp, and in the posterior temporal regions, the 'simultaneous' ERP was more positive in amplitude from about 45 to 80 ms compared to the 'sum' ERP, and had a smaller peak amplitude for the positive going wave at about 100 ms and the negative going wave at about 160 ms. At the central and fronto-central electrode sites the 'simultaneous' ERP was of greater amplitude than the 'sum' ERP for the negative going wave at about 120 ms, and for the positive going wave at about 180 ms.

3.2.2. Posterior AV interactions

The early right parieto-occipital difference between the 'simultaneous' and 'sum' ERPs was confirmed with a significant Stimulus Type by Hemisphere interaction (centered at 58 ms: F (1, 11) = 5.73, p = .036); the 'simultaneous' ERP was more positive than the 'sum' ERP (see Figure 4a). A follow up focused contrast over the right electrode sites, with factors of Stimulus Type and Electrode, showed a main effect of Stimulus Type (F (1, 11) = 8.83, p = .013).

In the latency range and scalp region of the occipitally focused visual P1, the 'simultaneous' ERP was significantly less positive going than the 'sum' ERP (centered at 120 ms: F (1, 11) = 5.89, p = .034). However, inspection of the voltage map suggested that the AV interaction in this latency range was largely focused over left centro-parietal scalp (see Figure 4c). A post-hoc two-way repeated measures ANOVA on left centro-parietally placed electrodes supported this contention (F (1, 11) = 15.69, p = .002) and we concluded that the AV effects in this latency range mainly reflected interactions over left centro-parietal scalp.

In the latency range and scalp area of the occipito-temporally focused visual N1 the 'simultaneous' ERP was less negative going than the 'sum' ERP (see Figure 4f), with a 2-way interaction of Stimulus Type by Electrode that approached significance (centered at 165 ms: F(1, 11) = 4.22, p = .061). The 'simultaneous' versus 'sum' difference was focused over right scalp. Given the strong trend toward significance we did a post hoc contrast on the right hemisphere electrode sites, with factors of Stimulus Type and Electrode; this showed a main effect of Stimulus Type (F (1, 11) = 6.31, p = .029.

3.2.3. Fronto-Central/Central AV interactions

In the latency range and scalp region of the fronto-centrally focused auditory N1, the amplitude of the 'simultaneous' ERP was significantly more negative going than the amplitude of the 'sum' ERP (centered at 120 ms: F (1, 11) = 10.20, p = 0.009). However, inspection of the topography of the effect suggested that it was largely due to the left centro-parietally focused AV interaction (see Figure 4c, and the results from the post-hoc ANOVA on left centro-parietal sites, centered at 120 ms, reported above).

In the latency range and scalp area of the centrally focused auditory P2 the 'simultaneous' ERP was significantly more positive going than the 'sum' ERP (see Figure 4g) (centered at 180 ms: F (1, 11) = 5.23, p = 0.043). Inspection of the simultaneous and sum ERPs over fronto-temporal scalp suggested that there were no amplitude differences focused over this area (centered at 150 ms: F (1, 11) = 1.03, p = .331).

3.2.4. Exploratory analysis of the spatio-temporal properties of the AV interactions

The p-values from the running t-test analysis comparing the amplitudes of the 'simultaneous' and 'sum' ERPs across the electrode montage and epoch provide a comprehensive picture of the spatio-temporal properties of the AV interactions (Figure

5). The earliest AV interaction onset at 46 ms, was evident only over the right parieto-occipital region of the scalp, and lasted for 20 ms (also see Figure 4a). This was followed by an AV interaction, not identified in our first phase of analysis, focused over frontal scalp that onset at 86 ms and lasted for up to 26 ms (also see Figure 4b). Next, there were simultaneous AV interactions focused over left centro-parietal and right posterior temporal regions of the scalp. Left centro-parietally it onset at 102 ms, and right posterior temporally it onset at 100 ms; these evolved into AV interactions focused over left centro-parietal scalp and right superior temporal scalp that peaked at 128 ms (also see Figures 4c, d, and e). These effects did not appear to be a modulation of the auditory N1. Starting at 160 ms there was an AV interaction focused over right occipito-temporal scalp that lasted until at least 180 ms, and this difference had a topography and time course consistent with the visual N1 (also see Figure 4f). Finally, there was a central AV interaction focused over left scalp that onset at 174 ms and extended out to 200 ms post stimulus onset (also see Figure 4g). We should note that the two latter AV interactions were present at only 3 to 4 electrode sites, while the other AV interactions were present at many electrode sites clustered within a scalp region.

3.2.5. The present multisensory effects are not an artifact of slow wave anticipatory potentials

The performance of a task in our experimental paradigm introduced the possibility of slow wave anticipatory potentials (e.g., the contingent negative variation; see 62). These slow potentials begin several hundred milliseconds prior to stimulus onset and continue for a short time thereafter. If present, anticipatory potentials would be represented twice in the 'sum' ERP and only once in the 'simultaneous' ERP, resulting in an artificial difference between the two. Problematically, this difference would appear to begin post-stimulus onset, due to baseline correction on the pre-stimulus portion of the ERPs, and would be indistinguishable from a genuine AV interaction (56).[1]

The current experiment was designed to minimize the incidence of anticipatory potentials. SOA was varied randomly between 750 and 3000 ms such that subjects could not accurately predict the arrival of the next stimulus, and hence would be discouraged from anticipating its onset (see 13). To confirm that anticipatory potentials did not contribute to the reported AV effects we compared the onset of our earliest effect for two different baseline criteria. If the effect was a consequence of anticipatory potentials, then its onset latency should shift for different baseline criteria (56). The early right parieto-occipital effect onset at the same latency of 46 ms for the two baseline criteria: 100 to 50 ms pre-stimulus onset; and our original 50 ms pre-stimulus to 10 ms post-stimulus onset. This analysis strongly suggests that anticipatory potentials did not contribute significantly to our AV effects.

To further rule out the contribution of slow wave anticipatory potentials to our AV interactions, we inspected the grand mean 'simultaneous' and 'sum' ERPs for differences in the pre-stimulus portion of the waves. The data were reprocessed (as described in the methods section) to look at a longer pre-stimulus epoch (250 ms pre-stimulus to 200 ms post-stimulus onset). No pre-stimulus differences were observed at anterior electrode sites, and only minimal pre-stimulus differences were observed at posterior electrode sites, again suggesting that anticipatory potentials did not contribute to our AV effects.

---

[1] We thank Dr. Steven Hillyard for pointing out this potential problem at the 2001 CNS conference.

3.3. C1 onset

The early AV interaction fell within the time range of when the initial response in primary visual cortices, indexed by C1, generally occurs (e.g. 10, 11, 21). We therefore set out to compare the onset of the early AV effect with the onset of C1. However, the C1 in the group averaged ERP to the visual alone stimulus condition was of relatively small amplitude, and activation did not reliably onset until the P1 portion of the ERP, at 76 ms (at occipital sites). The relatively small amplitude C1 was likely due to our use of multiple visual stimulus locations across subjects. That is, because of the anatomy of the retinotopically arranged human visual cortex, the scalp topography of C1 changes as a function of stimulus position (10). Thus, C1s to the same visual stimulus presented to different locations would have different scalp topographies, and when averaged together the C1 would appear diminished.

To get a more accurate measure of the onset of C1 to our visual stimulus, we presented it to a single location for an additional 7 subjects (mean age 25 ±7.9 SD; 5 female; 5 right-handed; one also participated in the main experiment). Only differences in procedure between this and the main experiment are noted. The stimulus was always presented to the lower right quadrant of the visual field (2.6 degrees lateral and 2.6 degrees vertical from fixation). To maintain arousal a simple oddball paradigm was used in which subjects made a right-handed button-press response to an infrequently occurring change in color (5% of trials). SOA varied randomly between 750 and 1000 ms. Stimuli were blocked into sequences of 100 trials, and the first 3 subjects completed 4 blocks of trials while the remaining subjects each completed 8 blocks of trials. The continuous EEG was divided into 100 ms pre-stimulus to 200 ms post-stimulus onset epochs. These were averaged for each subject to compute the ERP. Baseline was defined as the pre-stimulus epoch. The average number of accepted sweeps was 329 ±29 for the first 3 subjects and 721 ±21 for the remaining subjects.

The group averaged ERP exhibited a C1 that was largest over occipital scalp and appeared to diverge from baseline at ~40 ms. The t-test analysis showed that C1 onset at 44 ms over left occipital scalp and at 46 ms over central and right occipital scalp. Thus, with a temporal resolution of essentially 2 ms, the onset latencies of C1 and the early AV interaction were the same.

4. Discussion

The present study examined the spatial and temporal properties of cortical AV interactions to basic stimuli, while subjects performed a simple reaction-time task. In the behavioral data there was a significant reaction-time advantage when the visual and auditory stimuli were presented simultaneously compared to when they were presented alone - the so-called redundant sensory effect (RSE). There was substantial violation of the race model. Hence the RSE could not be accounted for by simple probability summation, providing strong evidence that multisensory neural interactions are responsible for this sizeable RSE.

4.1 The early AV effect

The electrophysiological data exhibited surprisingly early auditory-visual multisensory interactions, over right parieto-occipital scalp. The combination of the early timing of this effect and its scalp topography over early visual areas is consistent with modification of very early visual sensory processing by auditory inputs. As unisensory transmission to auditory cortex is considerably faster than it is to visual cortex, this suggested the following possible interpretation to us. First, auditory input activates primary auditory cortex within 15 ms of stimulus presentation and is then transmitted up the auditory processing stream. This input is then projected to "unisensory" visual areas through one of two possible pathways.

The first possibility is a direct, feedforward projection from primary auditory or auditory association cortices to early visual cortices. Until recently, it was largely believed that such connectivity did not exist. However, very recently, anatomic tracer studies have provided evidence for direct projections from both primary auditory cortex (A1) and auditory association areas to both V1 and V2 in macaques (18, 44). The second possibility is a more indirect pathway, in which feedforward auditory input reaches areas of AV multisensory convergence (e.g. the superior temporal polysensory region - STS), and is transmitted via feedback connections to earlier "unisensory" visual areas. The critical issue is one of timing. The question is whether there is sufficient time for auditory input to reach early visual areas to result in modulation of the later arriving visual input? Given the standard timing offset between the initial auditory and visual inputs to their respective primary cortices, there is a more than sufficient window of approximately 25-30 ms in which this process could occur (49).

4.1.1. Timing and what it tells us

The onset of an AV interaction at 46 ms in our study is much earlier in information processing than multisensory processing is generally assumed to occur, and reinforces the view that visual processing is modified by auditory inputs well before visual sensory analysis is complete. The C1 component of the VEP, our measure of initial input to visual cortices, and the AV effect, were found to onset essentially simultaneously. This timing of AV interactions relative to C1 onset suggests that the visual input to this multisensory process arrives through feedforward connections. Further, this earliest AV interaction has different scalp topography to that of the C1, with a focus that is more dorsal over right parieto-occipital scalp. Thus, on the basis of both timing and topography, we suggest that an early stage of the dorsal visual stream is the likely site of this AV interaction. Visual activation of such areas is essentially simultaneous with activation of V1 (46). The auditory input, on the other hand, conceivably arrived through *either* feedforward or feedback connections, as auditory cortices are activated by tones of moderate intensity by at most 15 ms (9).

This finding of an early AV effect contributes to recent evidence (20, 24, 47, 49) of multisensory interactions during early sensory processing. Models of sensory integration need to account for such early multisensory operations in addition to later multisensory processing that occurs following the extensive sensory analysis of the individual sensory inputs (33, 39). For example, based on functional imaging data, Calvert et al. (7) and Driver and Spence (16) suggest feedback accounts of multisensory processing in unisensory cortex, although they do not specify the point at which this occurs in information processing.

4.1.2. Candidate areas for initial integration of multisensory inputs

The right parieto-occipital focus of the earliest AV interaction is consistent with sources originating from one of several possible underlying brain areas: Regions of the visual motion processing system[2] (referred to here as MT+, which includes V5 and related areas), and known multisensory regions of the posterior parietal cortex and the superior temporal sulcus (STS). Of these three, STS is the most parsimonious with the detailed neurophysiology of visual processing in the macaque monkey. In the monkey, STS receives visual inputs only milliseconds after V1 (~3 ms; 46, 49, and there are cells in STS that receive converging input from auditory cortices (3, 5, 25). Further, human fMRI studies implicate STS in auditory-visual multisensory processing (e.g. 8, 41). However, while its auditory-visual response characteristics make STS a strong candidate for the source of the early AV interaction, the topography of the interaction observed here and in Giard and Perronet (24) seems posterior to what one might predict for a generator in STS.

MT+ is consistent with the topography of the early interaction, and there is strong evidence from monkey intracranial data that the arrival of the fastest inputs to V5 occurs at the same time as, and sometimes even earlier than, initial inputs to V1 (45, 46, 48); and human data have found similar timing (6, 22). Further supporting the possibility that

---

[2] Transient stationary stimuli, such as ours, activate the visual motion processing system (e.g., Schroeder, Mehta, & Givre, 1998).

MT+ contributed to the early AV interaction, multisensory processing in human MT+ for auditory-visual speech has been shown with fMRI (7; also see 4). Although monkey studies have failed to convincingly show that MT+ responds to auditory stimuli (e.g., unpublished data from our lab), it is possible that such multisensory cells in the macaque have not been sampled.

Also consistent with the topography of the early interaction is posterior parietal cortex, which is activated by a visual stimulus just a few milliseconds following V1 activation (46, 49). Human and monkey studies have shown that posterior parietal areas are involved in the multisensory representation of location and motion (2, 30). Further, monkey studies indicate that these areas include neurons that respond to both auditory and visual inputs early in information processing (e.g., 34; and see 2 for review). A caveat is that such convergence of auditory and visual inputs has only been found when monkeys have been extensively trained to make saccades to the auditory stimuli (25).

4.1.3. Functional significance of early AV integration

MT+ and the posterior parietal cortices are involved in the processing of motion, and are implicated in exogenously (stimulus) driven attentional orienting to moving stimuli. Consistent with the right hemisphericity of the early AV interaction, the exogenous attentional orienting systems in posterior parietal cortices are generally localized to the *right* hemisphere (e.g., 36, 42). The rapid detection and localization of moving objects is advantageous to survival, for example allowing one to dodge a hurtling object whistling through the air, and would be enhanced by the early integration of temporally coincident auditory and visual information in motion sensitive areas.

Behavioral studies have shown that detection and localization of visual objects are improved when there is a temporally coincident, location concordant, sound (52, 53).

4.2. The later AV effects

A centro-parietal AV interaction peaked at 120 ms. This interaction overlapped with the time course of both the auditory N1 and the visual P1. However, its centro-parietal topography suggested a different configuration of generators. This evolved into more lateralized activity focused over right superior temporal and left centro-parietal scalp that peaked at about 128 ms; these AV interactions were distinct from the activity in response to the auditory or visual alone stimulus conditions within the same latency range (see figure 4d for comparisons of the interactions with the activity to the auditory alone condition at the same latency).

A later AV interaction, over left central scalp (~180 ms), was consistent with generators in motor cortex and may reflect sensory motor integration. The timing of the effect is consistent with such an interpretation, with the mean RT to the simultaneously presented stimuli occurring ~75 ms later, as is the scalp topography since the right handed response used in this study would be expected to elicit a left lateralized motor response.

In addition to the above, post hoc analyses suggested two further AV interactions. One was over frontal scalp (~ 100 ms), and was unique to the 'simultaneous' stimulus condition (i.e. did not correspond to activity in response to either the visual or auditory alone stimulus conditions). Multisensory processing in the frontal cortices has been shown previously in both human and monkey studies (3, 8). The other was focused over right occipito-temporal areas (~ 165 ms). The timing and scalp topography of this effect was consistent with those of the visual N1, elicited by the visual alone stimulus condition. Auditory-based modulation of the visual N1 would suggest that auditory inputs can affect the visual object recognition system, as the visual N1 has been implicated in visual object recognition processes by many studies (e.g., 1, 14, 15).

5. Summary & Conclusions

Our electrophysiological data are in general agreement with those reported by Giard and Peronnet (24), the only comparable ERP study to date. Both studies revealed AV interactions of the same polarity over right parieto-occipital scalp (~ 40-50 ms), over occipito-temporal scalp (~ 165 ms), and over fronto-central (in Giard & Peronnet, 24) and central (the present study) scalp (~180 ms). However, our AV effects were of considerably shorter duration than those reported in Giard and Peronnet. Additionally, while we also found multisensory effects in the latency range of the auditory N1 and visual P1 components, high-density topographic mapping showed that these effects had different topographies than the unisensory components, suggesting different underlying source configurations for these multisensory effects. Furthermore, our data did not evidence fronto-temporal activity at about 150 ms as was found in Giard and Peronnet. Of course, due to the considerable differences in the paradigms used, it is not surprising that some differences were seen. Rather, given the general similarities in findings across these two studies, we suggest that AV multisensory processing comprises a base set of cortical interactions that will be present for a variety of stimuli and tasks.

These data contribute to the growing literature on the spatio-temporal properties of cortical multisensory processing in humans. The extent to which multisensory activity is unique to particular combinations of the senses versus common to various sensory combinations, and the impact of task effects on the observed activities, remains to be seen. For example, it has been well documented that there can be profound crossmodal influences on attentional orienting (e.g. 19, 23, 35, 51). Further, the current data highlight the importance of a detailed consideration of the timecourse of sensory processing within the individual sensory modalities. The inherent differences in transmission time to cortex of stimulation within the different sensory modalities will likely have considerable effects on the timecourse and potential neural areas at which multisensory interactions will occur.

Acknowledgments

Sincere appreciation to Ms. Beth Higgins and Ms. Deirdre Foxe for excellent technical assistance. Work supported in part by grants from the NIH -- NS30029-23 (WR) MH63434 (JJF) and MH61989 (CES).

References

[1] T. Allison, A. Puce, D. Spencer, G. McCarthy, Electrophysiological studies of human face perception I: Potentials generated in occipitotemporal cortex by face and non-face stimuli, Cereb. Cortex 9 (1999) 415-430.

[2] R.A. Andersen, L.H. Snyder, D.C. Bradley, J., Xing, Multimodal representation of space in the posterior parietal cortex and its use in planning movements, Annu. Rev. Neurosci. 20 (1997) 303-330.

[3] L.A. Benevento, J. Fallon, B. J. Davis, M. Rezak, Auditory--visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey, Exp. Neurol. 57 (1977) 849-872.

[4] R.A. Berman, C.L. Colby, Both auditory and visual attention modulate motion processing in area MT+, Cognit. Brain Res. (in press).

[5] C. Bruce, R. Desimone, C.G. Gross, Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque, J. Neurophysiol. 46 (1981) 369-384.

[6] H. Buchner, R. Gobbele, M. Wagner, M. Fuchs, T.D. Waberski, R. Beckmann, Fast visual evoked potential input into human area V5, Neuroreport 28 (1997) 2419-2422.

[7] G.A. Calvert, M.J. Brammer, E.T. Bullmore, R. Campbell, S.D. Iversen, A.S. David, Response amplification in sensory-specific cortices during crossmodal binding, Neuroreport 10 (1999) 2619-2623.

[8] Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection

of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect, NeuroImage, 14, 427-438.

[9] G.G. Celesia, F. Puletti, Auditory input to the human cortex during states of drowsiness and surgical anesthesia, Electroencephalogr. Clin. Neurophys. 31 (1971) 603-609.

[10] V.P. Clark, S. Fan, S.A. Hillyard, Identification of early visual evoked potential generators by retinotopic and topographic analyses, Hum. Brain Mapp. 2 (1995) 170-187.

[11] V.P. Clark, S.A. Hillyard, Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential, J. Cogn. Neurosci. 8 (1996) 387-402.

[12] R. Desimone, C.G. Gross, Visual areas in the temporal cortex of the macaque, Brain Res. 178 (1979) 363-380.

[13] E. Donchin, P. Tueting, W. Ritter, M. Kutas, E. Heffley, On the independence of the CNV and the P300 components of the human averaged evoked potential, Electroencephalogr. & Clin. Neurophys. 38 (1975) 449-461.

[14] G.M. Doniger, J.J. Foxe, M.M. Murray, B.A. Higgins, J.G. Snodgrass, C.E. Schroeder, D.C. Javitt, Activation time-course of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure processes, J. Cogn. Neurosci. 12 (2000) 615-621.

[15] G.M. Doniger, J.J. Foxe, C.E. Schroeder, M.M. Murray, B.A. Higgins, D.C. Javitt, Visual perceptual learning in human object based recognition areas: A

repetition priming study using high-density electrical mapping, Neuroimage 13 (2001) 305-313.

[16] J. Driver, C. Spence, Multisensory perception: beyond modularity and convergence, Curr. Biol. 10 (2000) R731-735.

[17] J.R. Duhamel, C.L.Colby, M.E. Goldberg, Congruent representation of visual and somatosensory space in single neurons of monkey ventral intraparietal cortex (VIP). In: J. Paillard (Ed.), Brain and Space, Oxford University press, 1991, pp. 223-236.

[18] A. Falchier, L. Renaud, P. Barone, H. Kennedy, Extensive projections from the primary auditory cortex and polysensory area stp to peripheral area v1 in the macaque, Soc. Neurosci. Abstr., Vol. 27, Program No. 511.21, 2001.

[19] J.J. Foxe, G.V. Simpson, S.P. Ahlfors, Cued shifts of intermodal attention: Parieto-occipital ~10 Hz activity reflects anticipatory state of visual attention Mechanisms, Neuroreport 9 (1998) 3929-3933.

[20] J.J. Foxe, I.A. Morocz, B.A. Higgins, M.M. Murray, D.C. Javitt, C.E. Schroeder, Multisensory auditory-somatosensory interactions in early cortical processing revealed by high density electrical mapping, Cognit. Brain Res. 10 (2000) 77-83.

[21] J.J. Foxe, G.V. Simpson, Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing, Exp. Brain Res. 142 (2002), 139-50.

[22] D.H. ffytche, C.N. Guy, S. Zeki, The parallel visual motion inputs into areas V1 and V5 of human cerebral cortex, Brain 118 (1995) 1375-1394.

[23] K.G. Fu, J.J. Foxe, M.M. Murray, B.A. Higgins, D.C. Javitt, C.E. Schroeder, Cross-modality cued attention-dependent suppression of distracter visual input indexed by anticipatory parieto-occipital alpha-band oscillations, Cognit. Brain Res. 12 (2001) 145-152.

[24] M.H. Giard, F. Peronnet, Auditory-visual integration during multimodal object recognition in humans: A behavioral and electrophysiological study, J. Cog. Neurosci. 11 (1999) 473-490.

[25] A. Grunewald, J.F. Linden, R.A. Andersen, Responses to auditory stimuli in macaque lateral intraparietal area I. Effects of training, Neurophysiology 82 (1999) 330-342.

[26] D. Guthrie, J.S. Buchwald, Significance testing of difference potentials. Psychophysiology 28 (1991) 240-244.

[27] K. Hikosaka, E. Iwai, H. Saito, K. Tanaka, Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey, J. Neurophysiol. 60 (1988) 1615-37.

[28] J. Hyvarinen, Y. Shelepin, Distribution of visual and somatic functions in the parietal associative area 7 of the monkey, Brain Res. 169 (1979) 561-564.

[29] R. Kinchla, Detecting target elements in multielement arrays: A confusability model, Percept. Psychophys. 15 (1974) 149-158.

[30] J.W. Lewis, M.S. Beauchamp, E.A. DeYoe, A comparison of visual and auditory motion processing in human cerebral cortex, Cereb. Cortex 10 (2000) 873-888.

[31] E. Macaluso, C.D. Frith, J. Driver, Modulation of human visual cortex by

crossmodal spatial attention, Science 289 (2000), 1206-1208.

[32] A. MacLeod, Q. Summerfield, A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use, Br. J. Audiol. 24 (1990) 29-43.

[33] D.W. Massaro, Speechreading: illusion or window into pattern recognition, Trends Cogn. Sci. 3 (1999) 310-317.

[34] P. Mazzoni, R.M. Bracewell, S. Barash, R.A. Andersen, Spatially tuned auditory responses in area LIP of macaques performing delaued memory saccades to acoustic targets, J. Neurophysiol. 75 (1996) 1233-1241.

[35] J.J. McDonald, L.M. Ward, Involuntary listening aids seeing: evidence from human electrophysiology, Psychol. Sci. 11 (2000) 167-171.

[36] M.M. Mesulam, A cortical network for directed attention and unilateral neglect, Ann. Neurol. 10 (1981) 309-325.

[37] J. Miller, Divided attention: evidence for coactivation with redundant signals, Cognit. Psychol. 14 (1982) 247-79.

[38] C. Miniussi, M. Girelli, C.A. Marzi, Neural site of the redundant target effect electrophysiological evidence, J. Cogn. Neurosci. 10 (1998) 216-230.

[39] J.R. Movellan, J.L. McClelland, The Morton-Massaro law of information integration: implications for models of perception, Psychol. Rev. 108 (2001) 113-48.

[40] M.M. Murray, J.J. Foxe, B.A. Higgins, D.C. Javitt, C.E. Schroeder, Visuo-spatial response interactions in early cortical processing during a simple reaction time

task: A high density electrical mapping study, Neuropsychologia 39 (2001) 828-844.

[41] I.R. Olson, J.C. Gatenby, J.C. Gore, A comparison of bound and unbound audio-visual information processing in the human cerebral cortex, Cognit. Brain Res. (this issue)

[42] J.V. Pardo, P.T. Fox, M.E. & Raichle, Localization of a human system for sustained attention by positron emission tomography, Nature 349 (1991) 61-64.

[43] T.W. Picton, S.A. Hillyard, H.I. Krausz, R. Galambos, Human auditory evoked potentials. I. Evaluation of components, Electroencephalogr. Clin. Neurophys. 36 (1974) 179-190.

[44] K.S. Rockland, H. Ojima, Calcarine area V1 as a multimodal convergence area, Soc. Neurosci. Abstr., Vol. 27, Program No. 511.20, 2001.

[45] M.T. Schmolesky, Y. Wang, D.P. Hanes, K.G. Thompson, S. Leutgeb, J.D. Schall, A.G. Leventhal, Signal timing across the macaque visual system. J. Neurophysiol. 76 (1998) 3272-3278.

[46] C.E. Schroeder, A.D. Mehta, S.J. Givre, A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque, Cereb. Cortex 8 (1998) 575-592.

[47] C.E. Schroeder, R.W. Lindsley, C. Specht, A. Marcovici, J.F. Smiley, D.C. Javitt, Somatosensory input to auditory association cortex in the macaque monkey, J. Neurophysiol. 85 (2001) 1322-1327.

[48] C.E. Schroeder, A.D. Mehta, J.J. Foxe, Determinants and mechanisms of attentional modulation of neural processing, Front. Biosci. 6 (2001) 672-

684.

[49] C.E. Schroeder, J.J. Foxe, The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex, Cognit. Brain Res. (in press).

[50] B. Selzer, D.N. Pandya, Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus monkey, Brain Res. 192 (1980) 339-351.

[51] C. Spence, J. Ranson, J. Driver, Cross-modal selective attention: on the difficulty of ignoring sounds at the locus of visual attention, Percept. Psychophys. 62 (2000) 410-424

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

[53] B.E. Stein, N. London, L.K. Wilkinson, D.D. Price, Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis, J. Cogn. Neurosci. 8 (1996) 497-506.

[54] B.E. Stein, M.A. Meredith, The Merging of the Senses, Cambridge, MA: MIT Press, 1993.

[55] S. Supek, C.J. Aine, D. Ranken, E. Best, E.R. Flynn, C.C. Wood, Single vs. paired visual stimulation: superposition of early neuromagnetic responses and retinotopy in extrastriate cortex in humans, Brain Res. 830 (1999) 43-55.

[56] W.A. Teder-Sälejärvi, J.J. McDonald, F. Di Russo, S.A. Hillyard, An analysis of audio-visual crossmodal integration by means of event-related

potential (ERP) recordings, Cognit. Brain Res. (in press).

[57] H.G. Vaughan, W. Ritter, The sources of auditory evoked responses recorded from the human scalp, Electroencephalogr. Clin. Neurophys. 28 (1970) 360-367.

[58] H.G. Vaughan, W. Ritter, R. Simson, Topographic analysis of auditory event-related potentials, Prog. Brain Res. 54 (1980) 279-285.

[59] H.G. Vaughan, J.C. Arezzo, The neural basis of event-related potentials. In: T.W. Picton (Ed.), Human event-related potentials. Handbook of Electroencephalography and Clinical Neurophysiology. Revised series, Vol. 3, Elsevier, Amsterdam, 1988, pp. 45-96.

[60] S.B. Vincent, The function of the vibrissae in the behavior of the white rat, Behavioral Monographs, 1, 1912.

[61] M.T. Wallace, M.A. Meredith, B.E. Stein, Integration of multiple sensory modalities in cat cortex, Exp. Brain Res. 91 (1992) 484-488.

[62] W.G. Walter, R. Cooper, V.J. Alderidge, W.C. McCallum, A.L. Winter, Contingent negative variation: an electric sign of sensorimotor association and expectancy in the human brain, Nature 203 (1964) 380-384.

Table 1. t-tests comparing the CP for the 'simultaneous' stimulus condition to the CP predicted by the Race Model over the first 5 quantiles, in which the 'simultaneous' CP exceeded the predicted CP.

Quantile	CP: simultaneous	CP: predicted	t_{11}	p <
1	.001	.000	1.49	.080
2	.002	.001	1.23	.120
3	.017	.007	2.76	.009
4	.118	.066	3.23	.004
5	.328	.253	2.61	.012

Figure 1. The 16 possible locations of the visual stimuli, with the number of subjects for whom a given location was used in the main experiment. The eccentricities of the stimulus locations with respect to central location were 1.56° for the upper left and right locations just above fixation, 2.33° for the higher upper left location, and 3.28° for the far lower left location. The far lower left location was used for all subjects in the follow up experiment, and had an eccentricity of 3.28°.

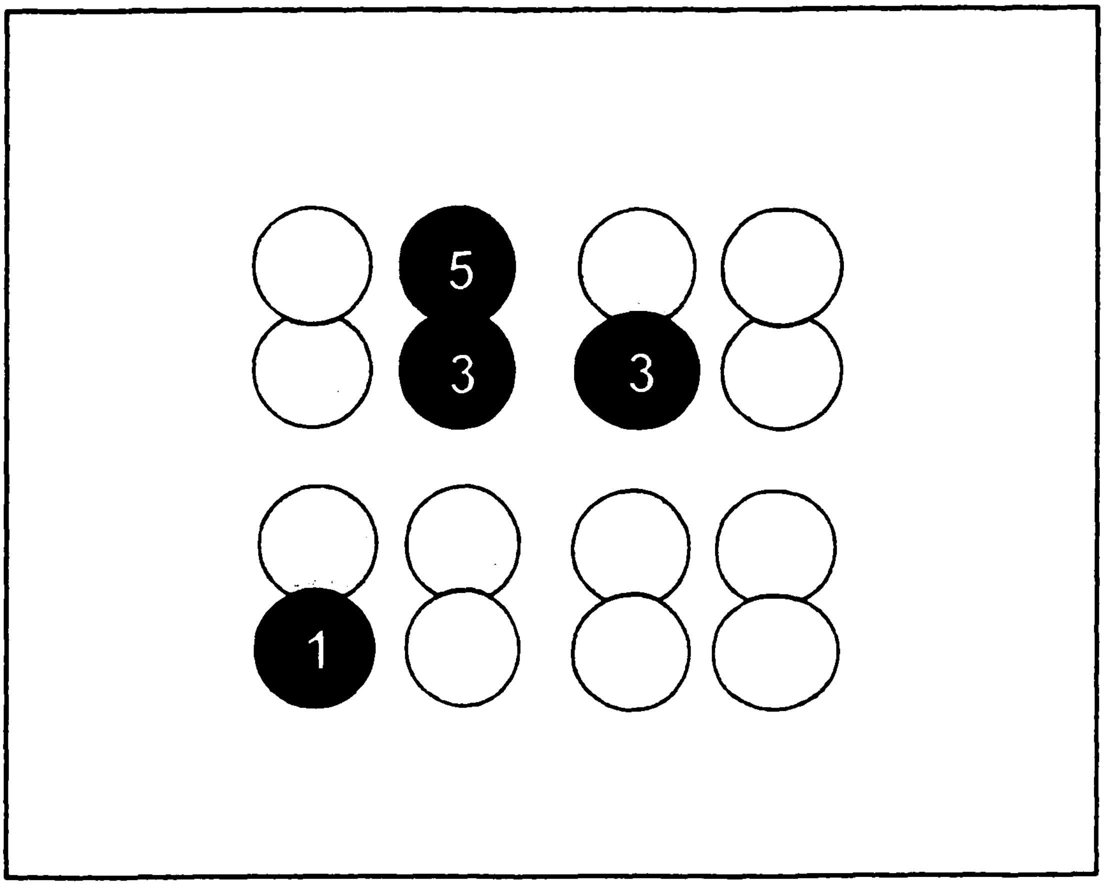

5
3
3
1

Figure 2. *a*. Mean (N=12) reaction times (SEM indicated) for simultaneous (AV: red bar), and auditory (A: blue bar) and visual (V: green bar) alone stimulus conditions. *b*. Cumulative probability (CP) distributions for the simultaneous (red trace) and visual (green trace) and auditory (blue trace) alone stimulus conditions, and the CP predicted by the race model (black trace). *c*. Miller inequality: values greater than zero signify violation of the race model (seen here over the first 6 quantiles).

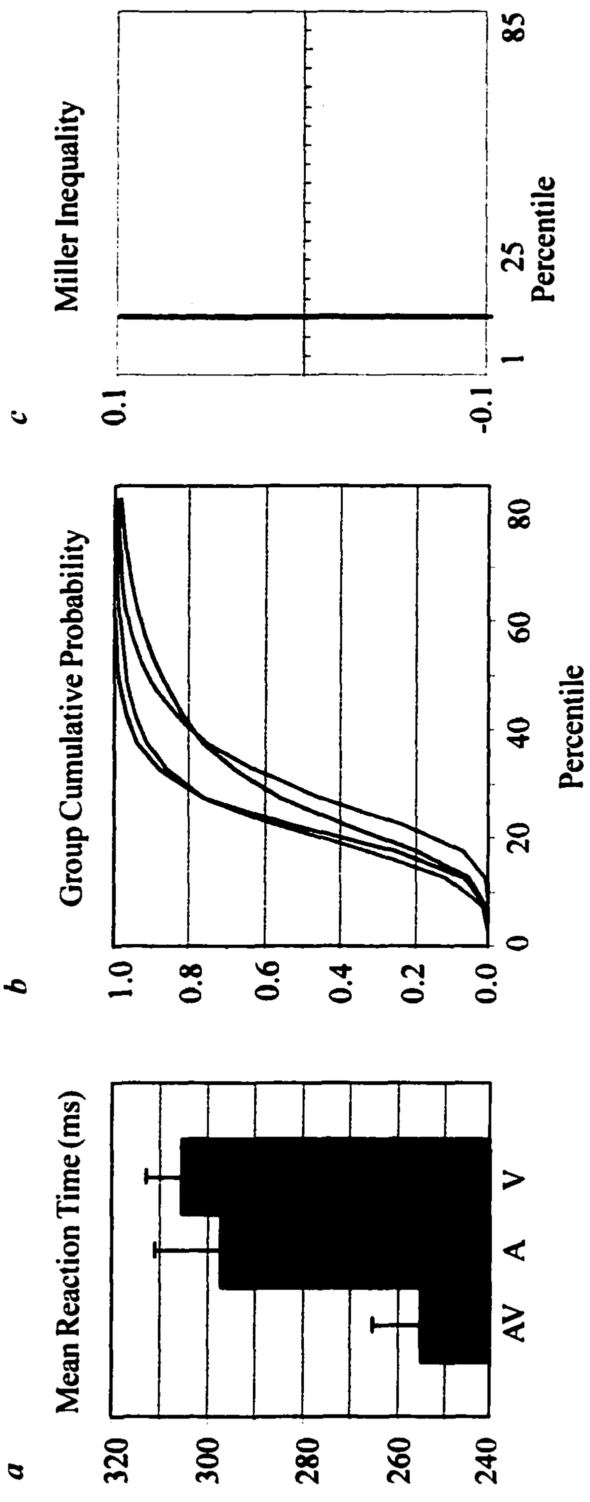

a
Mean Reaction Time (ms)
320
300
280
260
240
AV
A
V
b
Group Cumulative Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
Percentile
c
Miller Inequality
0.1
-0.1
1
25
85
Percentile

Figure 3. The 'simultaneous' (red trace), and visual (green trace) and auditory (blue trace) alone ERPs at left (*a*) and right (*b*) fronto-central electrode sites and at left (*c*) and right (*d*) occipital electrode sites. The placement of the electrode sites are indicated in magenta in top and back views of the electrode montage.

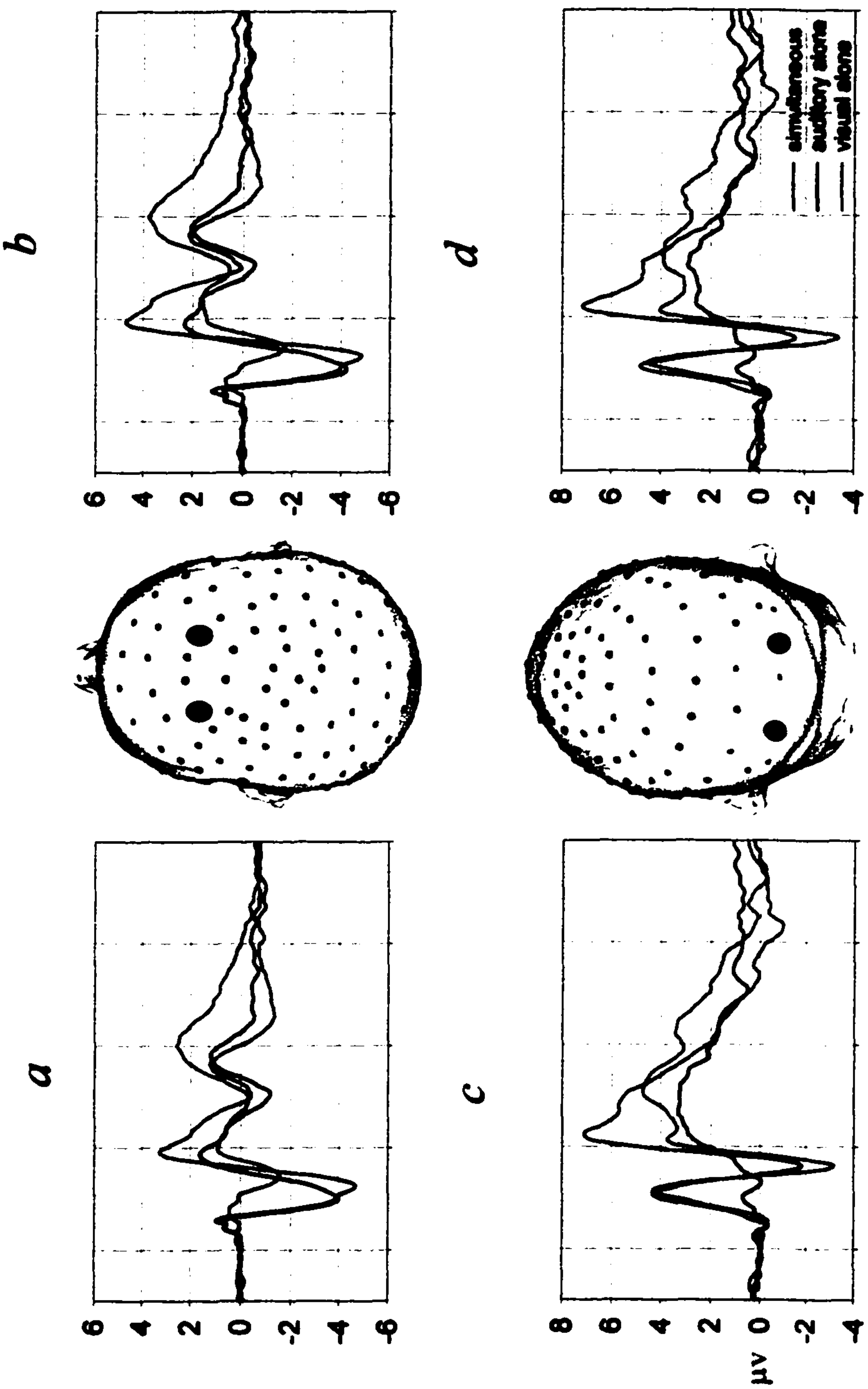
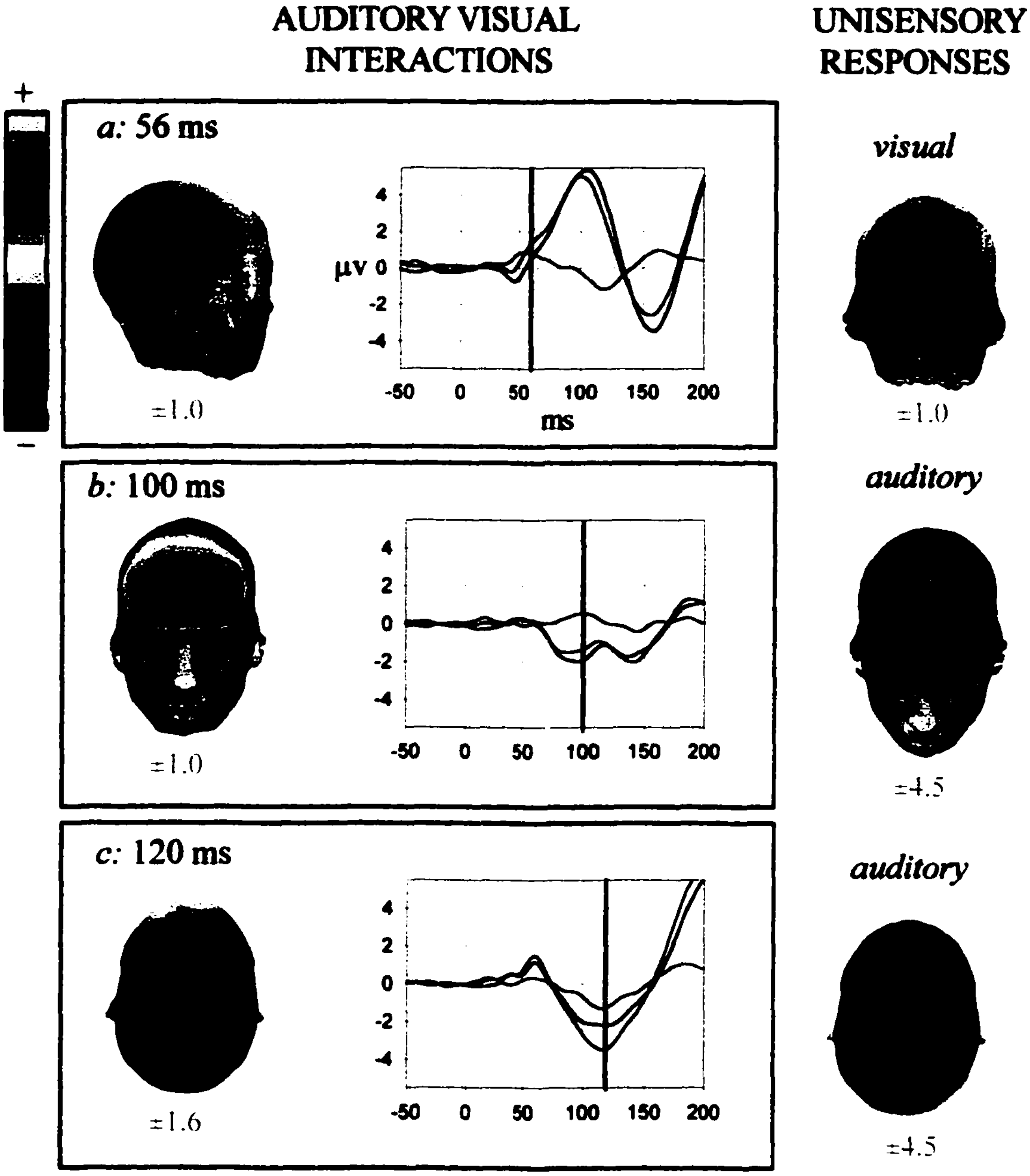

a
b
c
d
6
4
2
0
-2
-4
-6
8
6
4
2
0
-2
-4
µV
0
200
400
600
800
ms
simultaneous
auditory alone
visual alone

Figure 4. Auditory-visual interactions are illustrated in 'simultaneous' versus 'sum' difference voltage maps of the focus of the AV interactions, at their peak latencies. The color scale is to the left of Figure 4*a*, and the voltage scale used for each of the maps is indicated below it, in gray. These are accompanied by ERP traces to the 'simultaneous' (red), 'sum' (blue), and 'simultaneous' versus 'sum' difference (green) from corresponding electrode sites. The electrode location is indicated with a black dot on the voltage map. A vertical black line is drawn through the traces at the latency of the corresponding voltage map. For each AV interaction voltage map, a unisensory voltage map (visual or auditory) at the same latency is illustrated, to the far right. Comparison of the AV interactions and corresponding unisensory voltage maps reveals the similarity or difference of the interaction to unisensory processing.


AUDITORY VISUAL INTERACTIONS
UNISENSORY RESPONSES
+
−
a: 56 ms
±1.0
4
2
µV 0
-2
-4
-50
0
50
100
150
200
ms
visual
±1.0
b: 100 ms
±1.0
auditory
±4.5
c: 120 ms
±1.6
auditory
±4.5

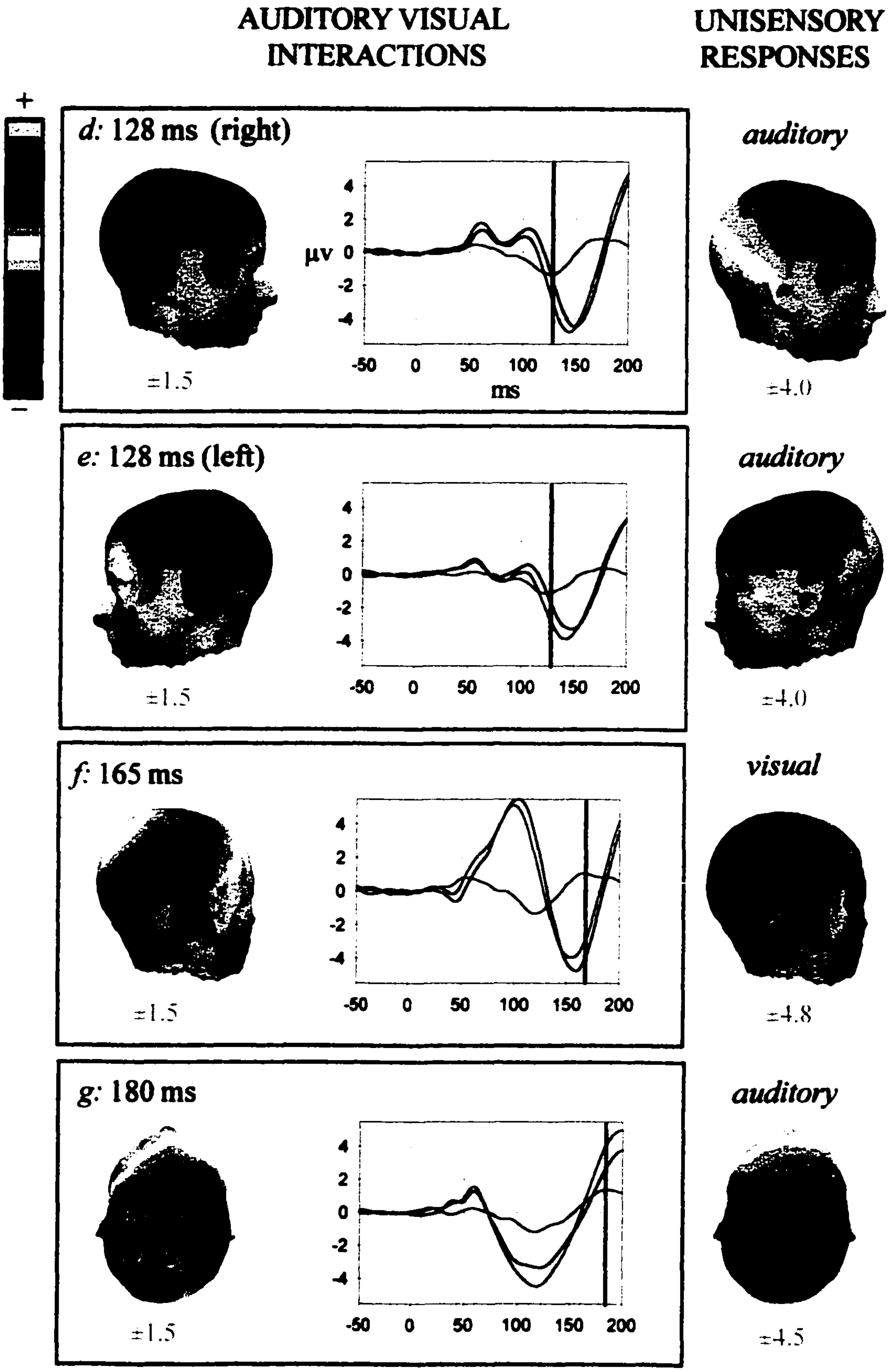
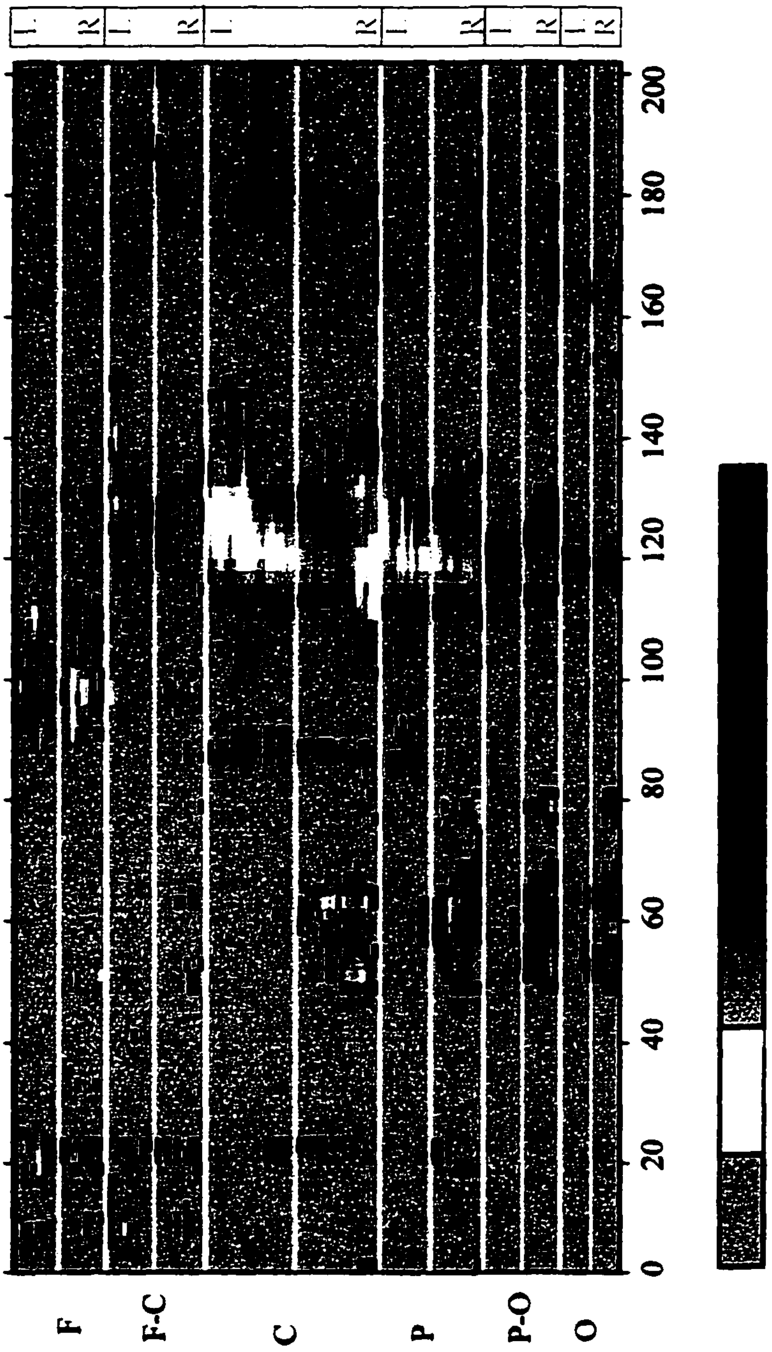

AUDITORY VISUAL INTERACTIONS
UNISENSORY RESPONSES
+
−
d: 128 ms (right)
4
2
µv 0
-2
-4
-50 0 50 100 150 200
ms
±1.5
auditory
±4.0
e: 128 ms (left)
4
2
0
-2
-4
-50 0 50 100 150 200
±1.5
auditory
±4.0
f: 165 ms
4
2
0
-2
-4
-50 0 50 100 150 200
±1.5
visual
±4.8
g: 180 ms
4
2
0
-2
-4
-50 0 50 100 150 200
±1.5
auditory
±4.5

Figure 5. Significant p values over time for 119 electrodes from running t-tests comparing the 'simultaneous' and 'sum' ERPs (this excluded 4 mastoid electrodes, 4 non scalp electrodes that recorded ocular activity, and 1 bad electrode) . P values are differentiated with a color scale. Grey indicates an absence of significant p values Time is plotted on the x axis from 0 to 200 ms. Electrodes are plotted on the y axis. Starting from the bottom of the graph, the electrodes are divided into sections from posterior to anterior scalp (O=occipital; P-O=parieto-occipital; P=parietal; C=central; F-C=fronto-central; F=frontal). Within a section the electrodes are arranged from the right most lateral to the left most lateral sites (indicated to the right of the graph for each section). The sections are labeled based on the mid line of the scalp (e.g., "central" includes electrodes over temporal scalp).



L
R
L
R
L
R
L
R
L
R
L
R
F
F-C
C
P
P-O
O
0
20
40
60
80
100
120
140
160
180
200
NS
p < .001
p = .01
p = .03
p = .05

CHAPTER 3

Multisensory visual-auditory object recognition in humans: A high-density electrical mapping study.

Abstract:

Integration of information from multiple senses is fundamental to perception and cognition, but when and where this is accomplished in the brain is not well understood. This study examined the timing and topography of cortical visual-auditory object recognition using high-density event-related potentials (ERPs) during an object recognition task. Visual and auditory elements were presented together that belonged to the same common object (congruent pairs) or belonged to different common objects (incongruent pairs). These were line drawings of animals and animal vocalizations from the same or different animals. Subjects responded to a target animal whether present in the visual, auditory, or both sensory modalities. We hypothesized that visual object recognition processes would be affected by the presence of ancillary auditory object information, and that this would result in reaction time facilitation for congruent targets, and modulation of the visual N1 ERP component. In support of this hypothesis, object recognition was speeded when the visual and auditory elements were object congruent. This RT facilitation could not be accounted for by probability summation, as evidenced

by violation of the 'race' model, indicating that visual-auditory interactions gave rise to this RT effect. In addition, we found that there was modulation of cortical processing for the same congruent target stimuli, in visual object recognition areas in the latency range of the visual N1, and over occipital scalp in the latency range of the selection negativity.

Introduction

Inputs to the various sensory systems function to inform us of the objects and events in the environment at any given moment, and our experience is one in which the multiple sensory properties cohere into unified perceptions of these object and events. To illustrate this, it helps to imagine what it is like when we are deprived of the contribution of one of these senses. For example, when it is dark we have to rely on auditory and somatosensory information in the near absence of visual information. In this case our experience is dramatically different, and objects and events can seem incomplete. The same would be the case were we deprived of auditory or somatosensory stimulation. Thus while we easily distinguish the unique characteristics of different sensory inputs (e.g., the touch, smell, and image of a flower), phenomenological experience, as well as experimental evidence (e.g., Stein, Huneycutt, & Meredith, 1988; Bertelson & Radeau, 1981; McGurk & MacDonald, 1976; McGrath & Summerfield, 1985), tells us that there are profound multisensory influences on the normal perception of objects and events.

Multiple sensory inputs that emanate from a single object usually provide redundant and/or complementary information about the identity of the object. For example, the smell and image of a rose are complementary information about the rose's identity; while the heard sound and seen speech articulations associated with a spoken syllable give partially redundant information about its identity. Oftentimes the information coming into the nervous system through any single sense is degraded due to "noise", such as the partial occlusion of an object or the hum of a vent masking someone's speech. In such instances sensory information that is complementary and redundant can be essential to accurate and timely object recognition.

Numerous studies have illustrated multisensory effects on perception, showing that viewing articulatory gestures can affect what speech sound is perceived (e.g., McGurk & MacDonald, 1976), or enhance speech perception in a noisy environment (Campbell & Dodd, 1980; MacLeod & Summerfield, 1990; Sumby & Polack, 1954; Thompson, 1995). However, multisensory object recognition has not been extensively examined outside the domain of speech perception. Related to multisensory object perception, several studies have considered multisensory influences on the perception of features. Jousmäki and Hari (1998) showed that tactile judgments of surface texture were influenced by simultaneous auditory stimulation, demonstrating that both tactile and auditory information can contribute to the perception of texture; while in another study, visual and tactile cues were both shown to contribute to the perception of height (Ernst & Banks, 2002).

The cortical spatio-temporal dynamics of the integration of object information from different senses, important for mapping out the sequence of neural processes that underlie multisensory object recognition, remains to be fully elucidated (for proposed models see Massaro 1999; Calvert 1998; Driver & Spence, 2000). Functional imaging and event related potential (ERP) studies of multisensory speech perception have detailed a network of cortical areas that plays a role in the integration of auditory-visual speech (e.g., Calvert, Brammer, Bullmore, et al, 1999; Calvert, Campbell & Brammer, 2000). However, it is probable that this circuit for speech perception is a relatively specialized one, and that the cortical network involved in integrating information from the different sensory systems depends on the class of multisensory objects/events that are being considered (e.g., animals versus speech), as well as the specific demands of the task that is being performed. Information regarding the timing of multisensory object processing relative to the well-detailed timing of ongoing unisensory processing is sparse at best (but see Giard & Peronnet, 1999, Molholm, et al., in press, and Schroeder & Foxe, in press).

The goal of the present study was to advance understanding of multiple sensory influences on object recognition. To this end we examined the combined influences of visual and auditory information on the recognition of animals, and placed these effects within the temporal and topographical framework of visual and auditory processing. High-density event-related potentials (ERPs) were recorded from 128 scalp electrodes while participants performed an object recognition task in which they pressed a key to the occurrence of a target animal. On each trial one of four stimulus types was presented: an animal picture (visual-unisensory); an animal sound (auditory-unisensory); a paired picture and sound of the same animal (object congruent-bisensory pair); and a paired picture of one animal and sound of another (object incongruent-bisensory pair). On a given run, one of eight possible animals was designated as the target. Five types of targets were derived from the four stimulus types: auditory-unisensory targets, visual-unisensory targets, auditory-incongruent-bisensory targets (with a target auditory element and a non-target visual element), visual-incongruent-bisensory targets (with a target visual elements and a non-target auditory element), and congruent-bisensory targets object congruent stimulus with target visual and auditory elements). Figure 1 illustrates the target and non-target stimuli, in the case where the target was a cow.

To assess the combined effect of visual and auditory information on object recognition, we compared performance on the bisensory targets to performance on the unisensory targets. We expected that the neural interaction of congruent visual and auditory object information (congruent-bisensory targets) would facilitate performance compared to the unisensory targets (auditory-unisensory targets and visual-unisensory targets). In contrast, we expected that the neural interaction of incongruent visual and auditory object information might interfere with recognizing the target (visual-incongruent-bisensory and auditory-incongruent-bisensory targets), and result in poorer performance compared to the unisensory targets. However, RT facilitation under these conditions could be accounted for by one of two classes of models: race models or coactivation models (see Miller, 1982, for a discussion of race and coactivation models). In race models each constituent of a pair of redundant targets independently competes for response initiation, and the faster of the two mediates the response for any given trial, resulting in the so-called redundant target effect (RTE). According to this model, probability summation produces the RTE, since the likelihood of either of the two targets yielding a fast RT is higher than that from one alone. In coactivation models, the interaction of neural responses to the simultaneously presented targets facilitates response initiation and produces the RTE. We therefore tested if the RTE exceeded the statistical facilitation predicted by the race model (Miller, 1982), and thereby provided evidence that object information from the visual and auditory sensory modalities interacted to produce RT facilitation - i.e. congruent-bisensory targets produced a violation of the race model.

Another question we asked was which were the likeliest brain structures to mediate these multisensory object recognition processes? Intuitively, object-recognition would be considered a visually dominated function. Consistent with this notion, there is data to suggest that even when an object is presented through a sensory modality other than vision, visual object recognition processes in the ventral-occipital stream may mediate object recognition. Up to this point this has only been established for stimuli presented tactilely (Amedi, Malach, Hendler, Peled, & Zohary, 2001; James, Humphrey, Gati, Servos, Menon, & Goodale, in press), not having been previously investigated for objects presented in other sensory modalities. However, our finding of the apparent auditory modulation of the visual N1 ERP in study 1 (chapter 2) suggested that this would also be the case for auditorily presented objects. Thus, we considered it most plausible that object-identity information provided in an adjunct sensory modality (in this case auditory) would have its effect in visual object-recognition areas, with the presence of features that defined an auditory object affecting processing of visual features of the same object.

The technique of high-density electrical mapping was used to establish the spatio-temporal dynamics of multisensory processing on object recognition. We predicted that the influence of ancillary auditory information on the recognition of animals would be reflected in the posterior visual N1 component of the ERP. The visual N1 occurs relatively early in information processing, typically reaching its peak amplitude between 140 and 170 ms. It is believed to be largely generated in structures of the ventral visual stream, as evidenced by intracranial grid-electrode recordings (e.g. Allison et al., 1999) and scalp topographic studies (e.g. Doniger, Foxe, Murray, et al, 2000; Doniger, Foxe, Schroeder, et al., 2001). The N1 (or one or more subcomponents of the N1 complex) has been repeatedly shown to reflect visual processing of the structural features of objects (e.g., Bentin et al., 1999; Doniger et al., 2000; 2001; Eimer, 2000; but see Vogel and Luck 2000) and thus to be involved in object-recognition processes. Using ERPs, Molholm and colleagues (2002) have shown visual-auditory multisensory interactions with a spatio-temporal profile similar to that of the visual N1 (see also Giard and Peronnet, 1999). Importantly, the requisite inter-sensory connectivity for auditory influences on visual processing has recently been described. Anatomic tracer studies have provided evidence for direct projections from primary auditory cortex (A1), auditory association cortex, and superior temporal polysensory cortex (STP; a multisensory convergence site) to both V1 and V2 in macaques (Rockland & Ojima, 2001; Falchier, Renaud, Barone, & Kennedy, 2001).

In designs where targets are defined by visual features such as shape or color, or auditory features such as frequency, well-characterized selective attention components are seen (Anllo-Vento & Hillyard, 1996; Harter & Aine, 1984; Hillyard, Hink, Schwent, & Picton, 1973; Kenemans, Kok, & Smulders, 1993; Näätänen, 1992; Näätänen, Gaillard, & Mäntysalo, 1978; Smid, Jakob, & Heinze, 1999). We were interested to see how these attentional components, which have been largely defined in studies that are purely unisensory or where visual and auditory stimuli have both been presented, but not simultaneously, would be affected by multisensory processing. We expected earlier effects of ancillary auditory inputs on processing in visual recognition areas (e.g., on the visual N1) to be passed on to later visual attentional components. In contrast, we did not expect visual inputs to affect auditory recognition processes; therefore we predicted that auditory attentional components would not be similarly affected.

Materials and Methods

Subjects

Fourteen neurologically normal, paid volunteers participated (mean age 23.6 ± 6.2; 6 female; all right-handed), all reported normal hearing and normal or corrected-to-normal vision. The Institutional Review Board of the Nathan Kline Institute for Psychiatric Research approved the experimental procedures, and each subject provided written informed consent. Data from two additional subjects were excluded, one for excessive blinking, and the other for failure to perform the task adequately.

Stimuli

There were four stimulus types: (1) animal sounds alone (auditory-unisensory); (2) animal pictures alone (visual-unisensory); (3) paired pictures and sounds of the same animal (congruent-bisensory pairs); (4) paired pictures of one animal and sounds of another animal (incongruent-bisensory pairs). In all there were 80 stimuli: 8 sounds, 8 pictures, and 64 pairings of each of the sound and picture stimuli.

Animal pictures: There were 8 line drawings (black on a gray background) of animals from the Snodgrass and Vanderwart set (1980), standardized on familiarity and complexity. These were of a dog, chimpanzee, cow, sheep, chicken, bird, cat, and frog.

They were presented on a monitor located 143 cm in front of the subject, and were black on a gray background. The images subtended an average of 4.8° of visual angel in the vertical plane and 4.4° of visual angle in the horizontal plane. These were presented for a duration of 340 ms.

Animal sounds: There were 8 complementary animal sounds, adapted from Fabiani, Kazmerski, Cycowicz, & Friedman (1996), of uniquely identifiable vocalizations corresponding to the 8 animal drawings. These were modified such that each had a duration of 340 ms, and were presented over 2 JBL speakers at a comfortable listening level of about 70 dB SPL.

Procedure

Participants were seated in a comfortable chair in a dimly lit and electrically shielded (Braden Shielding Systems) room and asked to keep head and eye movements to a minimum, while maintaining central fixation. Eye position was monitored with horizontal and vertical electro-oculogram (EOG) recordings. Subjects were first presented with the stimuli and asked to identify them. The sounds were presented first, followed by the pictures and finally the sound-picture pairs. Subjects easily identified the sounds and pictures as the animals intended by the experimenter.

During the experiment, subjects performed an animal detection task (e.g., "During this block, press the button to the cow") and withheld responses to any other animal. Button presses were made with the right index finger. A target was defined by the visual sensory modality, the auditory sensory modality, or both. There were 5 types of targets: visual-unisensory target (e.g., a picture of a cow), auditory-unisensory target (e.g., a "moo" of a cow), a picture and sound pair in which only the picture was the target (visual-incongruent-bisensory target; e.g., a picture of a cow and a bark of a dog), a picture and sound pair in which only the sound was the target (auditory-incongruent-bisensory target; e.g., a picture of a chimpanzee and a moo of a cow), and an object congruent picture and sound target pair (congruent-bisensory target; e.g., a picture of a cow and a moo of a cow). Each animal served as the target in 2 of 16 blocks, in randomized order both within and between subjects, with the exception that the full set of animals served as targets within the first 8 blocks. Targets occurred on 15.6% of trials within a block. There were 4 types of non-targets: an animal picture (visual-unisensory); an animal sound (auditory-unisensory); a paired picture and sound of the same animal (congruent-bisensory pair); and a paired picture of one non-target animal and sound of another non-target animal (incongruent-bisensory pair).

Each block contained 56 instances of each stimulus type such that a full set of incongruent stimulus pairs was presented in each block (8 visual animals X 7 incongruent auditory animal sounds). There were 7 tokens of each of the 5 target types in each block for a total of 35 targets in a block. Stimulus onset asynchrony varied randomly between 750 and 3000 ms. Sixteen blocks were presented, for a total of 896 trials for each of the 4 types of stimuli. Breaks were encouraged between blocks to maintain high concentration and prevent fatigue.

Data acquisition and analysis

Continuous EEG was acquired from 128 scalp electrodes (impedances < 5 kΩ), referenced to the nose, band-pass filtered from 0.05 to 100 Hz, and digitized at 500 Hz.

The continuous EEG was divided into epochs (-100 ms pre- to 500 ms post- stimulus onset) and baseline corrected over the full 600 ms. Trials with blinks and eye movements were automatically rejected off-line on the basis of the EOG. An artifact criterion of ± 60 µV was used at all other scalp sites to reject trials with excessive EMG or other noise transients. The average number of accepted sweeps per non-target was 670, and per target was 95. EEG epochs were sorted according to stimulus type and averaged from each subject to compute the event-related potential (ERP). Baseline was defined as the epoch from –100 ms to stimulus onset. Separate group-averaged ERPs for each of the stimulus types were calculated for display purposes and for identification of the occipito-temporal N1.

Button press responses to the 5 target stimuli were acquired during the recording of the EEG and processed offline. Responses falling between 250 and 950 ms post stimulus onset were considered valid. This window was used so that a response could only be associated with a single trial.

Statistical analyses

Behavioral.

For individual subjects, the percent hits and average response-time (RT) were calculated, and reaction-time distributions were recorded, for each of the 5 target stimuli. To test for an effect of Target on RT, a one-way repeated measures analysis of variance (ANOVA) was performed with 5 levels (auditory-unisensory target, visual-unisensory target, auditory-incongruent-bisensory target, visual-incongruent-bisensory target, and congruent-bisensory target). Tukey HSD tests were performed to determine which of the

Target RTs differed significantly. The same statistical analyses were performed to test for an effect of Target on percent hits.

Significant RT facilitation for the congruent targets was followed by two tests of Miller's test of the race model. This was done to determine if race or coactivation models best accounted for the RTE. In one test of the race model, the congruent-bisensory targets were compared to the auditory-incongruent-bisensory and visual-incongruent-bisensory targets; and in the second test of the race model, the congruent-bisensory targets were compared to the visual-unisensory and the auditory-unisensory targets. We assumed that if violations of the race model were due to interactions between the visual and auditory object information, they should be present in both of these tests. This is because in each of the tests of the race model, the congruent targets differed from the comparison targets in that they had simultaneously presented object-congruent visual and auditory target elements. See Figure 2 for an illustration of this point.

The race model places an upper limit on the cumulative probability (CP) of RT at a given latency for stimulus pairs with redundant targets. For any latency, t, the race model holds when this CP value is less than or equal to the sum of the CP from each of the single target stimuli (either 'unisensory-targets' or 'incongruent targets'; the formulas are written for the test with the 'unisensory-targets') minus an expression of their joint probability $(CP_{(t)\text{congruent}} \leq ((CP_{(t)\text{A-unisensory}} + CP_{(t)\text{V-unisensory}}) - (CP_{(t)\text{A-unisensory}} * CP_{(t)\text{V-unisensory}}))$. For each subject the RT range within the valid RTs (250-950 ms) was calculated over the three target types (congruent-bisensory, visual-unisensory, and auditory-unisensory targets; or congruent-bisensory, visual-incongruent, and auditory incongruent targets) and divided into quantiles from the fifth to the hundredth percentile in 5% increments (5%, 10%,..., 95%, 100%). T-tests comparing the actual facilitation ($CP_{(t)\text{congruent-Bi}}$) and facilitation predicted by the race model (($CP_{(t)\text{A-unisensory}} + CP_{(t)\text{V-unisensory}}$) - ($CP_{(t)\text{A-unisensory}} * CP_{(t)\text{V-unisensory}}$)) were performed on quantiles that exhibited violation of the race model, to assess the reliability of the violations across subjects. Violations were expected to occur for the quantiles representing the lower end of the RTs, because this is when it was most likely that interactions of the visual and auditory inputs would result in the fulfillment of a response criterion before either source alone satisfied the same criterion (Miller, 1982).

Interference of target detection by the simultaneous presentation of visual and auditory elements that did not belong to the same object, on the auditory-incongruent-bisensory and visual-incongruent-bisensory target trials, was indexed by a significant increase in RT or decrease in percent hits for the auditory- and visual-incongruent-bisensory targets over, respectively, the auditory- and visual-unisensory targets, as tested with the Tukey HSD tests.

Event-related potentials

For the ERP analyses only the bisensory stimuli were considered. This is because the visual and auditory sensory ERP componentry was matched across the bisensory stimuli. Recall that the possible bisensory combinations under the current design were: 1) congruent bisensory targets (e.g., a picture of a cow and the moo of a cow); 2) visual-incongruent-bisensory targets (e.g., a picture of a cow and the bark of a dog); 3) auditory-incongruent targets (e.g., a picture of a chimpanzee and the moo of a cow); 4) congruent-bisensory non-targets (a picture of a dog and the bark of a dog); and 5) incongruent-bisensory non-targets (a picture of a chicken and the croak of a frog).

Visual-auditory congruency effects over posterior scalp in the latency range of the visual N1

We had a specific hypothesis that visual object recognition processes in the ventral visual stream would be affected by ancillary auditory information. We therefore tested for visual-auditory object congruency effects on the mean amplitude of the ERPs over a 20 ms window, centered at the peak of the N1. These were submitted to a three-way repeated measures ANOVA with factors of Stimulus (5: congruent-bisensory non-target, incongruent-bisensory non-target, congruent-bisensory target, visual-incongruent-bisensory target, and auditory-incongruent-bisensory target), Electrode (3 for each hemisphere), and Hemisphere (left and right). Follow up ANOVAs were conducted to unravel the congruency effects. For all statistical tests Geisser-Greenhouse corrections were used in reporting *p* values when appropriate and the alpha level for significance was less than .05.

Multisensory effects on selective processing components

We hypothesized that visual selective attention components would be affected by ancillary auditory inputs, while auditory selective attention components would not be affected by visual inputs. As such we examined the selection negativity (SN), a negative going potential over occipital scalp, and the selection positivity (SP), a positive going potential over frontal scalp, which are elicited by relevant visual stimuli when relevance is defined on the basis of a non-spatial feature(s) (Anllo-Vento & Hillyard, 1996; Harter & Aine, 1984; Kenemans, Kok, & Smulders, 1993; Smid, Jakob, & Heinze, 1999), and the negative difference wave (Nd), which is elicited by relevant auditory stimuli when relevance is defined on some physical dimension(s) (e.g., Hillyard, Hink, Schwent, & Picton, 1973; Näätänen, 1992; Näätänen, Gaillard, & Mäntysalo, 1978). We expected the bisensory non-target ERPs to have similar morphologies in the latency ranges and scalp areas of these selective attention effects. Therefore for each of the following tests, they were collapsed after establishing that there was no significant difference between the congruent-bisensory and incongruent-bisensory non-targets. The presence of a significant difference a three way ANOVA with factors of non-target stimulus (2: congruent and incongruent bisensory non-target ERPs), hemisphere (2: right and left), and electrode (3) was performed.

The selection positivity

To test for the SP we first performed a three-way repeated measures ANOVA on the mean ERP amplitudes over frontal scalp, for a 100 ms window starting at the point where the ERPs diverged on the basis of the target status of their visual element. This had factors of Stimulus (5: congruent-bisensory targets, visual-incongruent bisensory target, auditory-incongruent bisensory target, congruent bisensory non-target, and incongruent bisensory non-target), Hemisphere (right and left), and Electrode (3). A significant effect of Stimulus was followed up with three three-way repeated measures ANOVAs, with factors of Stimulus (2: in each of the tests, one of the targets was contrasted with the collapsed non-target), Hemisphere (right and left), and Electrode (3). These served to compare each of the target ERPs to the collapsed non-target ERP to determine if they differed significantly. We expected that each of the target ERPs with a visual target element (congruent bisensory targets and visual-incongruent bisensory targets) would be significantly more positive going than the collapsed non-target ERP, while the target ERP without the 'relevant' visual target element would not be significantly more positive going than the collapsed non-target ERP. To test for the effect of ancillary auditory information on the SP, the congruent-bisensory target ERP, which had visual and auditory elements that were targets, and the visual-incongruent-bisensory target ERP, which only had a visual element that was a target, were similarly compared.

The selection negativity

To test for the SN, the same statistical tests were performed as for the 'selection positivity' (see sub-section above), on the mean ERP amplitudes over lateral occipital scalp, for a 30 ms window starting at the point where the ERPs diverged on the basis of the target status of their visual element. We expected each of the ERPs elicited by a stimulus with a target visual element to be significantly more negative going than the collapsed non-target ERP, while the target ERP elicited by the stimulus without a target visual element would not be significantly more negative going. To test for the effect of ancillary auditory information on the SP we tested for a significantly enhanced SN in the congruent-bisensory target ERP, which had visual and auditory elements that were targets, compared to the visual-incongruent-bisensory target ERP, which only had a visual element that was a target with a three-way ANOVA with factors of Stimulus (2: congruent-bisensory target and visual-incongruent bisensory target), Hemisphere (2: right and left), and Electrode (3).

The negative difference wave

To test for the Nd, we first performed a three-way repeated measures ANOVA on the mean ERP amplitudes over central scalp, for a 30 ms window starting at the point where the ERPs diverged on the basis of the target status of their auditory element. This had factors of bisensory stimulus (5: congruent-bisensory targets, visual-incongruent bisensory target, auditory-incongruent bisensory target, congruent bisensory non-target, and incongruent bisensory non-target), hemisphere (right and left), and electrode (3). A significant effect of stimulus was followed up with three three-way repeated measures ANOVAs, with factors of Stimulus (2: in each of the tests, one of the targets was contrasted with the collapsed non-target), Hemisphere (right and left), and Electrode (3). We expected that each of the target ERPs with an auditory target element (congruent bisensory targets and auditory-incongruent bisensory targets) would be significantly more negative going than the collapsed non-target ERP, while the target ERP without the 'relevant' auditory target element would not be significantly more negative going than the collapsed non-target. To test for an enhancement of the Nd due to congruency, the congruent-bisensory target ERP and the auditory-incongruent-bisensory target ERP were similarly compared.

Results

Behavioral

In the object recognition task, both reaction times and percent hits were affected by target type (see Table 1). The pattern of results strongly suggests that object recognition was enhanced by the co-occurrence of object-congruent visual and auditory information. Object recognition for the congruent-bisensory targets was superior to object recognition for the other targets, with the fastest mean reaction time and highest mean hit rate. Visually based object recognition was better and faster than auditory based object recognition, with faster mean reaction times and higher mean percents hits for targets that included a visual target component. Overall, the false alarm rate was low at 2%; false alarms were more or less evenly distributed among the auditory-unisensory, visual-unisensory, incongruent-bisensory and congruent-bisensory non-targets.

Reaction times

Mean reaction times were substantially different among the Targets (see Table 1), with the difference between the shortest and longest reaction time greater than 130 milliseconds. A main effect of Target ($F(4, 52) = 90.16$, $p < .001$) was followed up with Tukey HSD comparisons. These revealed significant RT differences between all the target stimuli except the visual-unisensory target and the visual-incongruent-bisensory target, and the auditory-unisensory target and the auditory-incongruent-bisensory target (see Table 2). The mean RT to the congruent-bisensory targets was significantly faster than the mean RTs to all the other targets, demonstrating that targets were more rapidly recognized in the congruent objects. The comparisons also revealed that the RTs to the targets with a visual element that was a target were significantly faster than the RTs to the target stimuli with an auditory element that was a target (excluding the congruent-bisensory target).

Test of the Race Model

The race model was reliably violated for 4 successive quantiles in the early portion of the reaction time distribution (the 3rd through the 6th: see Table 3 and Figure 3).

This was the case both when RT performance for congruent targets was compared to the incongruent-bisensory targets (see Figure 2- test 1), and when RT performance for the congruent targets was compared to the unisensory targets (see Figure 2- test 2), strongly suggesting that the interaction of visual and auditory object information during neural processing contributed to the RTE (see Figure 3).

Percent hits

Percent hits also differed considerably among the targets, with a difference of about 15% between the highest and the lowest hit rates. The pattern of differences generally paralleled those of the reaction times. For example, by both measures, the best performance was for the congruent-bisensory targets and the worst performance was for the auditory-incongruent-bisensory targets. There was a main effect of Target type (F (4, 52) = 17.96, p < .001). Tukey HSD comparisons revealed that percent hits were significantly higher for stimuli that included a visual target element compared to trials that included an auditory target element, excluding the congruent-bisensory targets (see Table 2).

Interference effects for incongruent-bisensory targets were not apparent in our data, with a lack of significant differences between the auditory-incongruent-bisensory targets and auditory-unisensory targets and the visual-incongruent-bisensory targets and visual-unisensory targets, by both the RT and percent hit measures.

Event-related potentials

The group averaged ERPs elicited by the 5 bisensory stimuli are displayed in Figure 4. The ERPs showed similar classic auditory and visual sensory componentry (see figure 4). These included typical auditory components P1 peaking at ~60 ms and N1 peaking at ~118 ms (Picton, Hillyard, Krausz, & Galambos, 1974; Vaughan & Ritter, 1970; Vaughan, Ritter, & Simson, 1980); and visual components P1 peaking at ~78 and N1 peaking at ~150 ms. The auditory P1 and N1 appeared maximal over fronto-central scalp. The visual P1 appeared maximal over lateral occipital scalp and the visual N1 appeared maximal over posterior-temporal scalp (see Figure 5, bottom). While the P1 and N1 are usually more posterior (Clark et al., 1995; Foxe & Simpson, 2002), a similar posterior-temporal N1 topography has been reported in response to faces (e.g., Carmel & Bentin, 2002; Eimer, 2000) and other objects such as houses (e.g., Eimer, 2000; Rossion, Gauthier, Tarr, et al., 2000).

The ERPs to all the bisensory stimuli, whether they were congruent or incongruent, were essentially identical until ~125 ms post-stimulus onset. At this point, over orbito-frontal scalp, the ERP elicited by the auditory-incongruent-bisensory stimuli diverged from the other ERPs in the negative direction (see Figure 6, bottom panel). This was followed in time by the divergence of the ERP elicited by the congruent-bisensory stimuli from the other ERPs, over lateral occipital scalp- this difference was present from ~145 to 165 ms post-stimulus onset, falling within the latency range of the visual N1, but with a scalp distribution posterior to the posterio-temporally focused N1 (see Figure 5). Next, over central/centro-parietal scalp, the ERPs elicited by stimuli that included a target auditory element, i.e., the congruent-bisensory targets and auditory-incongruent-bisensory targets, became more negative going than the ERPs elicited by the stimuli that did not include a target auditory element, i.e., visual-incongruent-bisensory targets, congruent-bisensory non-targets and incongruent-bisensory non-targets (see Figure 6, top panel). This 'auditory target effect' was apparent starting at about 180 ms, and suggested that there were cortical systems that processed the auditory elements of the stimuli as a function of their relevance. Shortly after the onset of the centrally focused effect, the ERPs elicited by the targets with a target visual element (congruent-bisensory targets and visual-incongruent-bisensory targets) became more positive going than the ERPs elicited by the non-targets (congruent- and incongruent-bisensory non-targets) and the target without a visual element (auditory-incongruent-bisensory targets), over orbito-frontal scalp (see Figure 6, middle panel). This difference began at about 190 ms, and appeared as a slow wave shift that persisted throughout the epoch. This effect suggested that there were cortical systems that processed the visual elements of the stimuli as a function of their relevance. Starting at ~200 ms, the ERPs elicited by the targets were more negative going than the ERPs elicited by the non-targets over occipital scalp, this difference appeared slightly larger over the left hemisphere. This negativity was considerably enhanced in the ERP elicited by the congruent-bisensory targets, starting at ~210 ms (see Figure 7, top and middle panels). This enhancement suggested that when both the visual and auditory elements of a stimulus were targets, they interacted in visual cortex to facilitate selective processing of the relevant stimuli whether defined by their visual or auditory elements, or possibly target specific processing. Finally, the ERP elicited by the incongruent non-targets was more negative than the ERP elicited by the congruent non-targets starting at about 390 ms (see figure 7, bottom panel), suggestive of late onsetting object congruency effects for the non-targets.

The effect of 'object congruency' on visual processing in the latency range of the N1

Over right occipito-temporal scalp the congruent-target ERP appeared to be of greater amplitude and duration than the ERPs elicited by the other stimulus types in the latency range of the visual N1, with the remaining 4 ERPs close to superimposed. The overall ANOVA showed a Stimulus by Hemisphere interaction (F (4, 52) = 3.75, p = .02). Three follow up three-way ANOVAs with factors of Congruency (2), Electrode (3) and Hemisphere (2) were conducted to unravel the effects of object congruency on the 'N1'. Each of the ANOVAs compared the ERPs evoked by two of the stimulus types, one 'object congruent' and the other 'object incongruent' (congruent-bisensory target versus visual-incongruent-bisensory target; congruent-bisensory target versus auditory-incongruent-bisensory target; and congruent-bisensory non-target versus the incongruent-bisensory non-target). The congruent-bisensory target and visual-incongruent-bisensory target comparison revealed a significant Congruency by Hemisphere interaction (F (1, 13) = 8.08, p = .01; see Figure 5, middle panel). The interaction was due to a larger effect of Congruency over the right hemisphere. The congruent-bisensory target and auditory-incongruent-bisensory target comparison revealed a main effect of Congruency (F (1, 13) = 8.7, p = .01; see Figure 5, upper panel). The congruent-bisensory non-target and incongruent-bisensory non-target comparison failed to reveal any significant effects (for the main effect of Congruency, (F (1, 13) = 1.56, p = .23)). Thus there was an effect of congruency, and this was specific to the targets.

Visual selective processing over frontal scalp: the selection positivity

The ERPs exhibited a selection positivity, with the stimuli with a target visual element more positive going than the stimuli without a target visual element starting at 210 ms (see Figure 6, middle panel, in which the average of the bisensory ERPs with a target visual element (the congruent-bisensory and visual-incongruent-bisensory target ERPs) is compared to the average of the bisensory ERPs without a target visual element (auditory-incongruent bisensory target and non-target ERPs). Therefore the mean ERP amplitude from 210 to 310 ms was used for these tests. The overall ANOVA resulted in a main effect of Stimulus (F (4, 52) = 11.3, p = .00). Since there was no significant difference between the non-targets (F (1, 13) = 4.2, p = .527) we collapsed them for the follow up tests. There were significant differences between each of the three bisensory target ERPs and the collapsed non-target ERP (for congruent-, visual-incongruent-, and auditory-incongruent-bisensory ERPs respectively, F (1, 13) = 7.96, p = .01; F (1, 13) = 10.6, p = .006; and F (1, 13 = 22.9, p = .000)). The first two were more positive going than the non-target ERP, while the auditory-incongruent-bisensory ERP was more negative going. This supported that there was selective processing of relevant visual elements, and indicated that ancillary auditory information did not interact with this effect.

Visual selective processing over occipital scalp: the selection negativity

We expected selective processing of visual targets to be reflected over lateral occipital scalp, and that this would be affected by ancillary auditory inputs. The bisensory target ERPs appeared more negative going than the bisensory non-target ERPs starting at 210 ms, and this was considerably enhanced in the congruent bisensory target ERP. Therefore the mean ERP amplitude from 210 to 240 ms was used for these tests. The overall ANOVA revealed a main effect of Stimulus (F (4, 52) = 11.8, p = .000). Since there was no significant difference between the non-targets (F (1, 13) = 0.25, p = .62) we collapsed them for the follow up tests. The congruent-bisensory target ERP was significantly more negative going than the collapsed bisensory non-target ERP, with a main effect of Stimulus (F (1, 13) = 56.0, p = .000), this effect interacted with Hemisphere (F (2, 26) = 6.6, p = .03) due to a greater difference for the electrodes positioned over the left scalp. The visual-incongruent-bisensory target ERP was significantly more negative than the collapsed bisensory non-target ERP over the left scalp, with a Stimulus by Hemisphere interaction (F (2, 26) = 10.6, p = .006). We were surprised that the auditory-incongruent bisensory target ERP was negative going with respect to the collapsed bisensory non-target ERP, appearing quite similar in morphology in this latency window over lateral occipital scalp to the visual-incongruent bisensory target ERP. In line with these observations, the auditory-incongruent-bisensory target ERP versus the collapsed bisensory non-target ERP comparison approached significance (F (1, 13) = 4.4, p = .05). There was a significant difference between the congruent-bisensory target ERP and the visual-incongruent-bisensory target ERP (F (1, 13) = 16.10, p = .001), indicating that ancillary auditory information indeed increased the amplitude of the SN (see Figure 7 voltage maps, top and middle panels, in which the congruent-bisensory target ERP is subtracted from the average of the ERPs without a target visual element (top panel: auditory-incongruent bisensory target, congruent-bisensory non-target and incongruent-bisensory non-target ERPs) and the average of the incongruent-bisensory target ERPs (middle panel).

Auditory selective processing over central scalp: The negative difference wave

The ERPs elicited by stimuli with an auditory element that was a target (congruent-bisensory targets and auditory-incongruent-bisensory targets) appeared more negative going than the ERPs elicited by stimuli with an auditory element that was not a target (visual-incongruent-bisensory targets, congruent-bisensory non-targets, and incongruent-bisensory non-targets) starting at 190 ms, with a focused over left central scalp (see Figure 6, top panel). The overall ANOVA revealed a main effect of Stimulus (F (4, 52) = 4.89, p = .004). Since there was no significant difference between the non-targets (F (1, 13) = 1.45, p = .25) we collapsed them for the follow up tests. There was a significant difference between both of the congruent-bisensory and auditory-incongruent bisensory target ERPs and the collapsed bisensory non-target ERP (respectively, F (1, 13) = 7.34, p = .008 and F (1, 13) = 7.88, p = .015), while the visual-incongruent bisensory target did not differ significantly from the collapsed bisensory non-target ERP (F (1, 13) = 0.01, p = .923). There was no evidence of an effect of visual information on this auditory effect, with no significant difference between the congruent- and auditory-incongruent-bisensory target ERPs (F (1, 13) = 0.4, p = .53).

Exploratory analyses

We conducted additional tests in a 'post-hoc' phase of the analysis. These were based on our observations of the ERP waveforms and voltage maps. They were performed to broaden our description of the data, and had the added benefit of providing data to be used as a hypothesis generation tool for future multisensory studies.

A late occurring congruency effect in the non-target stimuli

There was a late occurring congruency effect specific to the non-targets (see Figure 7, bottom panel). A two way ANOVA with factors of Congruency (2: congruent and incongruent-bisensory non-targets) and Electrode (3) was performed on the mean ERP amplitudes over a 100 ms window (400-500) from 3 electrodes over central scalp where the difference was greatest. A main effect of Congruency (F (1, 13) = 9.68, p = .008) strongly suggested an object congruency effect on the non-target bisensory ERP.

Multisensory interactions in the congruent-bisensory targets

The two earlier object congruency effects were specific to the ERPs elicited by stimuli that were targets. We tested for evidence of multisensory neural interactions for the congruent and incongruent-bisensory targets in the latency ranges of these congruency effects, with the hypothesis that they would be specific to (or larger for) the congruent-bisensory targets. Multisensory interactions were assessed by comparing the bisensory target ERP to the summed ERPs of the visual- and auditory-unisensory targets (hereafter summed ERP); differences signified that the neural responses differed when the visual and auditory elements were presented simultaneously compared to when they were presented alone (see Figure 8; for a full rational of this method of measuring multisensory neural interactions see Chapter 2).

Each of the ERPs elicited by the 3-bisensory targets (congruent, visual-incongruent, and auditory-incongruent) was compared to the summed ERP in the latency ranges where we had confirmed the earlier object congruency target effects. These tests were performed on the sets of electrodes over the hemisphere where these congruency effects had been maximal (3 right occipital for the first effect and 3 left occipital for the second effect). For the latency window of 140-160 ms, three two-way ANOVAs with factors of ERP (2: Bisensory versus Summed) and Electrode (3) were performed. There was a main effect of ERP when the congruent-bisensory target ERP was compared to the summed ERP (F (1, 13) = 5.33, p = .038); this was due to the greater negativity of the bisensory ERP. The bisensory and summed ERPs did not differ significantly between either of the incongruent-bisensory target ERPs and the summed ERP (F (1, 13) = .712, p = .414 and F (1, 13) = .32, p = .57, respectively for each of the visual and auditory-incongruent-bisensory target ERPs compared to the summed ERP). Similarly, for the latency window of 210-240 ms, three two-way ANOVAs with factors of ERP (2: Bisensory versus Summed) and Electrode (3) were performed. There was a main effect of ERP when the congruent-bisensory target ERP was compared to the summed ERP (F (1, 13) = 17.37, p = .001); this was due to the greater negativity of the bisensory ERP. The bisensory and summed ERPs did not differ significantly between either of the incongruent-bisensory target ERPs and the summed ERP (F (1, 13) = .1.5, p = .23 and F (1, 13) = .84, p = .37, respectively for each of the for visual and auditory-incongruent-bisensory target ERPs compared to the summed ERP). These results strongly suggest that neural multisensory interactions in these latency ranges occurred for the congruent-bisensory targets but not for the incongruent-bisensory targets.

Discussion

The purpose of this study was to examine the cortical spatio-temporal properties of multisensory visual-auditory object recognition processes. Our main goal was to test the prediction that ancillary information in the auditory-sensory modality would modify processing in visual object recognition areas of the ventral visual stream. The data strongly suggested that this was the case, with adjunct auditory information modifying processing over occipital scalp. The target specificity of the object congruency effect suggested that it might be due to the simultaneous presence of task relevant information in each of the visual and auditory sensory modalities, as opposed to the semantically based congruency of the visual and auditory elements of the object. Thus under the conditions of this experiment, auditory effects on visual processing at the featural level appeared to be task dependent. The electrophysiological effects were accompanied by enhanced object recognition for the same stimulus type (i.e., congruent-bisensory targets). Violation of the race model indicated that the neural interaction of visual and auditory object information contributed to this enhanced behavioral performance.

ERP effects focused over occipital scalp are usually associated with the processing of strictly visual information (but see Molholm et al., 2002 and Schroeder et al, 2002 for a discussion of this issue), however recent evidence indicates that auditory influences on visual processes in occipital cortex are entirely feasible. Direct anatomical connections from auditory to visual areas have been identified in macaques (Falchier et al., 2001; Rockland & Ojima, 2001), and two ERP studies have shown visual-auditory multisensory interactions over occipito-temporal cortex in the latency range of the visual N1 (Giard & Peronnet, 1999; Molholm et al., 2002).

An alternative explanation of the target specificity of the object congruency effect is that stimulus elements were processed selectively based on their target status, and that cognitively costly operations were only performed on the elements that were targets. For example, subjects could have compared visual and auditory inputs to visual and auditory sensory representations of the targets (e.g., a line drawing of a cow and a moo of a cow), and terminated processing on the non-target elements once the absence of a relevant visual or auditory feature (or the presence of a non-target feature) was detected. Termination of the processing of irrelevant stimuli could explain why there were only object congruency effects for the targets, which were ‘relevant’. By this explanation, the multisensory effects were due to object congruency, and would be observed in non-targets under appropriate conditions. Problematically, there was a congruency effect in the non-target ERPs, which occurred *after* the termination of processing would have had to take place by this explanation

Selective processing of the 'relevant' (i.e., target) visual and auditory elements of the stimuli

We predicted that visual selective processing would be enhanced by ancillary auditory information, while the selective processing of auditory information would not be similarly affected by visual information. The electrophysiological data supported this contention, pointing to the selective processing of the visual and auditory target elements by distinct cortical systems, along with a cortical system over occipital scalp that uniquely responded supra-additively when both visual and auditory target information was presented. That is, the congruent-bisensory target ERP was of greater amplitude than the summed ERP (i.e., the sum of the ERPs elicited by the visual-unisensory targets and the auditory-unisensory targets).

Auditory influences on the selective visual processing

Over occipital cortex, the ERPs elicited by the targets were more negative going than the ERPs elicited by the non-targets. This negativity was considerably enhanced in the ERP elicited by the congruent-bisensory targets. This enhancement suggested that when both the visual and auditory elements of a stimulus were targets, they interacted in visual cortex to facilitate the processing of the relevant target stimuli. We were surprised that the auditory-incongruent-bisensory target ERP exhibited a 'selection negativity' that approached significance. This suggested to us that visually dominated object recognition processes were affected by relevant auditory information even in the absence of relevant visual information, under the conditions of this experiment.

The selection positivity over frontal cortex, on the other hand, was not affected by ancillary auditory information. The selective processing was specific to the visual sensory modality, with the ERPs differentiating largely based on the target status of the visual element of the stimulus[3].

Auditory target effect

Stimuli with an auditory element that was a target elicited a more negative going ERP than stimuli with an auditory element that was not a target over central/centro-parietal scalp. Consistent with our hypothesis that multisensory effects on object processing would take place in visual object recognition areas, this was not affected by the concurrent presence of congruent visual information.[4]

---

[3] We do not know why the erp elicited by the auditory-incongruent-bisensory target was more negative going over frontal scalp than the ERPs elicited by the other stimuli. One possibility is that the 'no-go' potential (Sasaki & Gemba, 1986; Gemba & Sasaki, 1989; Sasaki & Gemba, 1989; Sasaki, Gemba, Nambu, & Matsuzaki, 1993) also contributed to the frontally focused ERP evoked by the auditory-incongruent-bisensory targets. The no-go potential is a frontal slow wave negativity that is attributed to the inhibition of a planned motor response (Sasaki & Gemba, 1986; Gemba & Sasaki, 1989; Sasaki & Gemba, 1989). It could be the case that the target auditory element of the stimulus resulted in the planning of a motor response, which was then inhibited by the lack of a target visual element in the stimulus. This however seems unlikely, for one because there was no concomitant difference in RTs between auditory-incongruent-bisensory targets and auditory-unisensory targets, the latter of which would not have elicited a like no-go response.

[4] We should point out that the conditions under which the Nd is usually observed differ substantially from those of the present study. The Nd is typically examined by subtracting the ERP elicited by a non-target stimulus when it is 'relevant' from the ERP elicited by the same non-target stimulus when it is 'irrelevant', in a selective attention task in which subjects attend to a specified stream of auditory events that occur on 50% of the stimulus trials (with the irrelevant stimuli occurring on the other 50% of trials).

Conclusions

These data strongly suggest that when visual and auditory features were task relevant they interacted relatively early in information processing, and as a consequence facilitated the recognition of stimuli with task relevant visual and auditory elements (i.e., congruent targets). Based on the scalp topography and coincident timing with visual object-recognition processes, these proposed feature-based interactions most likely occurred in posterior visual cortical areas, and involved the influence of ancillary auditory information on predominantly visual processes. Later on in time stimuli were processed differentially based on the semantic congruity of their visual and auditory elements. This suggests a two stage (at the least) model of the higher order integration of visual and auditory information- the first of which occurs at the featural level and is task dependent, and the second of which occurs at the semantic level and is not task dependent.

We propose that target selection processes were achieved through the matching (and mismatching) of input to featural representations of the targets, which could be quickly achieved based on the presence or absence of particular feature; whereas the

---

Thus it is entirely possible that what we were looking at was not a traditional auditory selective attention effect, but rather a target effect specific to the auditory sensory modality. On the other hand, similar to the present study, the Nd has been elicited by relevant stimuli that occur at a relatively low probability (16.7% in Alho, Lavikainen, Reinikainen, Sams, & Naatanen, 1990; and 9% here), and it has been observed in ERPs elicited by targets compared to ERPs elicited by different non-target stimuli (Woods & Alain, 2001). One reason to favor an Nd interpretation is that inspection of the ERP difference wave derived from the auditory-unisensory target ERP minus the auditory-unisensory non-target ERP suggested an Nd with the same time-course. This had a fronto-central distribution more typical of the Nd that is elicited when only auditory stimuli are presented in the course of the experimental run. However, this target versus selective attention distinction is not vital here. In either case the differential processing of 'relevant' auditory elements was indexed.

subjects waited until they could identify the presence of the target animal in the visual and or auditory elements of a stimulus before responding, in which case the dynamic nature of the auditory information prolonged identification of the auditory elements relative to the static visual elements (To identify auditory inputs, often 100 ms of the stimulus or more is required (e.g., Grosjean, 1980; 1985); whereas with a line drawing, all the visual inputs needed to identify it are available at time zero). The non-target ERPs suggested a visual-auditory object congruency effect, which was apparent starting at about 380 ms. The late latency of this object congruency effect suggests that the integration of object information (as opposed to featural information) from the visual and auditory sensory modalities occurred after sensory specific perceptual analysis was complete, either during or after semantic processing. We assume that object congruency also affected processing of the targets, but that such effects were obscured by target related potentials.

Table 1.

	Congruent	Incongruent visual	Visual alone	Auditory alone	Incongruent auditory
RTs in ms	492 (75)	531 (68)	530 (58)	610 (53)	624 (53)
% hits	96 (4)	95 (5)	92 (17)	84 (12)	80 (17)

Table 2.

Reaction Time Differences (upper right) & Percent Hit Differences (lower left)

	Congruent	Incongruent visual	Visual alone	Auditory alone	Incongruent auditory
Congruent		38.20 **	37.60 **	117.57 **	131.86 **
Incongruent-visual	1.50 --		0.57 --	79.36 **	93.64 **
Visual alone	4.59 --	3.06 NS		79.90 **	94.20 **
Auditory alone	12.67 --	11.10 **	8.08 *		14.26 --
Incongruent-auditory	16.60 **	15.13 **	12.07 **	3.97 --	

** significant at $p < .01$
* significant at $p < .05$

Table 3. t-tests comparing the CP for the congruent-bisensory targets to the CP predicted by the Race Model for test 1 and test 2 over the first 6 quantiles.

Test 1

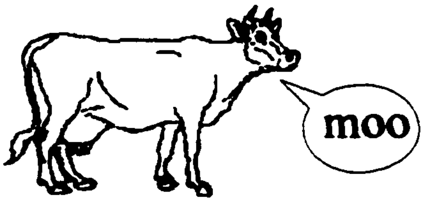
Quantile	CP: congruent	CP: predicted	t_{11}	p =
1	.006	.006	0.53	.410
2	.002	.004	1.39	.060
3	.019	.011	1.90	.038
4	.050	.044	2.17	.024
5	.117	.092	4.33	.000
6	.173	.120	4.70	.000

Test 2

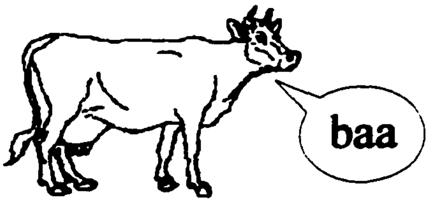
Quantile	CP: congruent	CP: predicted	t_{11}	p =
1	.001	.002	0.30	.380
2	.019	.012	0.96	.170
3	.051	.036	2.25	.020
4	.116	.075	3.50	.002
5	.159	.136	3.55	.002
6	.147	.178	3.18	.004

Figure 1. Examples of the 5 target (left side) and 4 non-target (right side) stimulus types are illustrated, in the case where the target was a cow

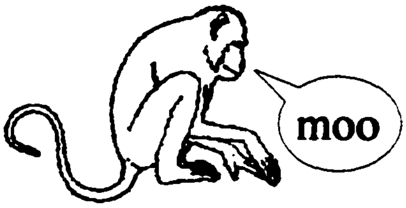
TARGETS (COW)



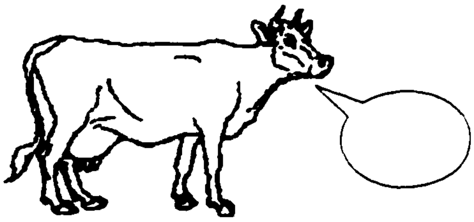
congruent bisensory target



visual incongruent bisensory target



auditory incongruent bisensory target

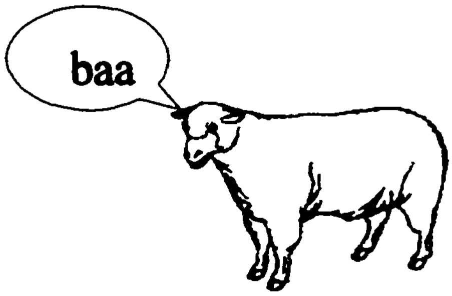


visual unisensory target

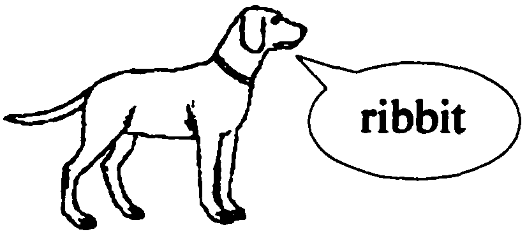
"moo"

auditory unisensory target

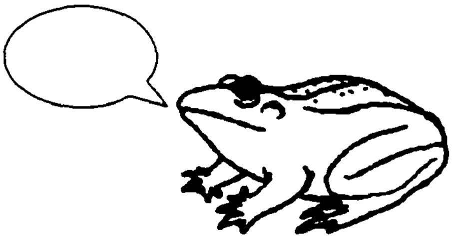
NON-TARGETS



congruent bisensory non-target



incongruent bisensory non-target



visual unisensory non-target

"woof-woof"

auditory unisensory non-target

Figure 2. This illustration is a schematic of the two tests of the race model. This makes the point that for both test 1 and test 2 of the race model, RT facilitation for trials on which both visual and auditory elements of the stimulus were targets were compared to predicted RT performance in cases where the visual and auditory target elements were presented separately.

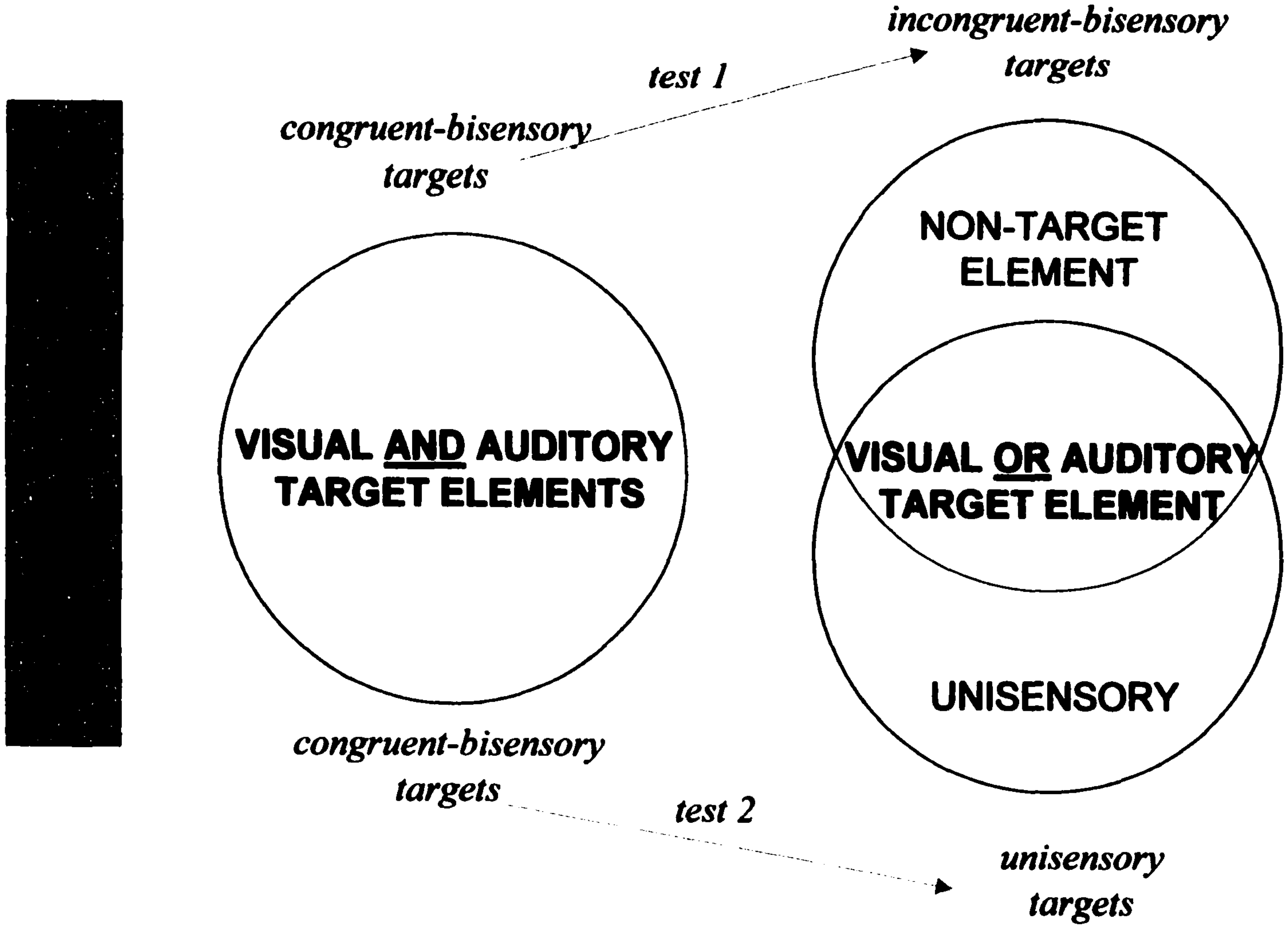

test 1
incongruent-bisensory targets
congruent-bisensory targets
NON-TARGET ELEMENT
VISUAL AND AUDITORY TARGET ELEMENTS
VISUAL OR AUDITORY TARGET ELEMENT
UNISENSORY
congruent-bisensory targets
test 2
unisensory targets

Figure 3. The left most column graphs reaction-time performance for the congruent-bisensory targets compared to the incongruent-bisensory targets (test1) and compared to the unisensory targets (test 2). The middle column shows the cumulative probability of reaction-times from the 5th to the 85th percentile, for congruent-bisensory targets (red curve), visual-incongruent-bisensory targets (or visual-unisensory targets- green curve), auditory-incongruent-bisensory targets (or auditory-unisensory targets- blue curve), and the cumulative probability predicted by the race model (yellow dashed curve). The right most column graphs the Miller Inequality: positive values indicate violation of the race model (i.e., the cumulative probability of responses for the congruent-bisensory targets was greater than the cumulative probability predicted by the race model, at the specified percentile).

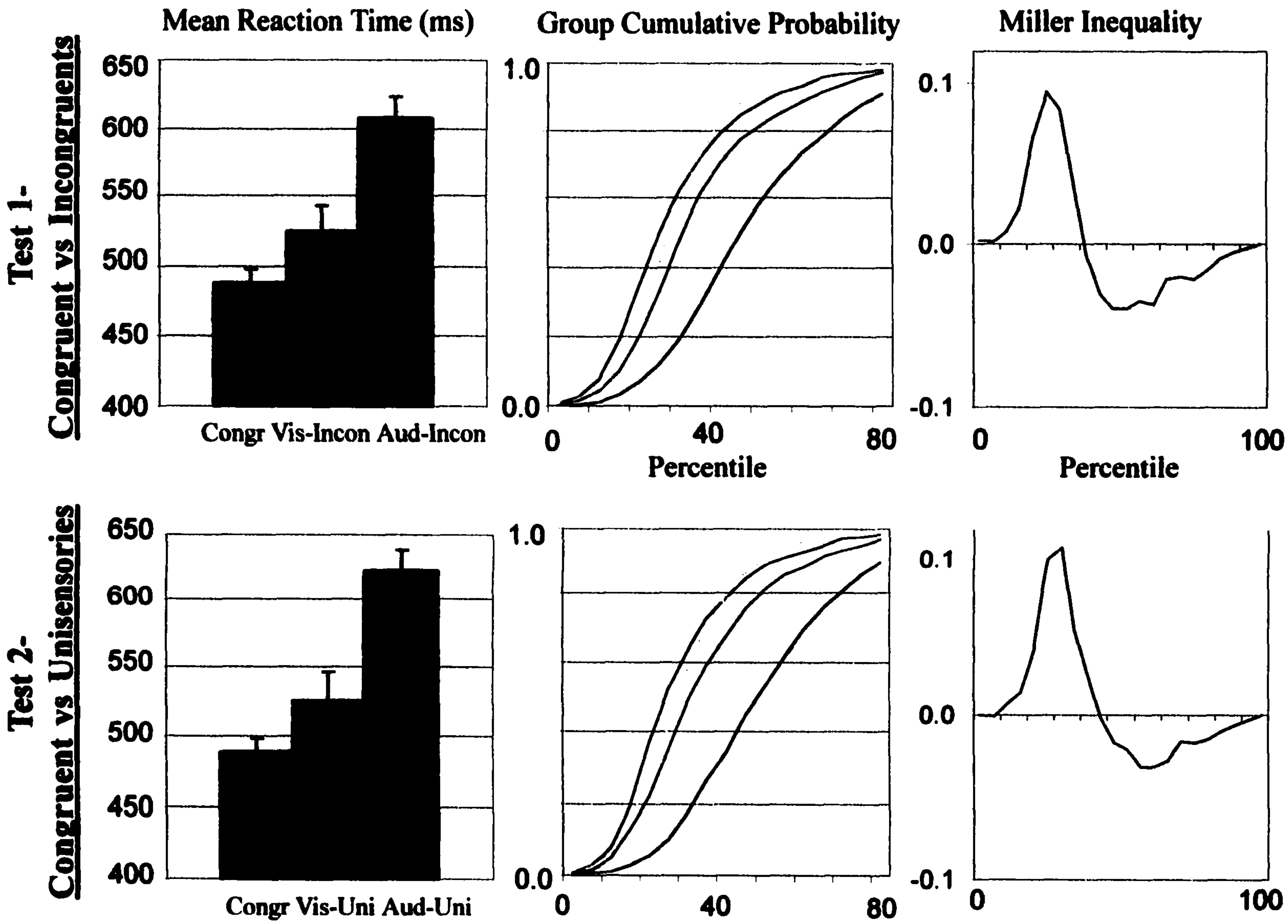
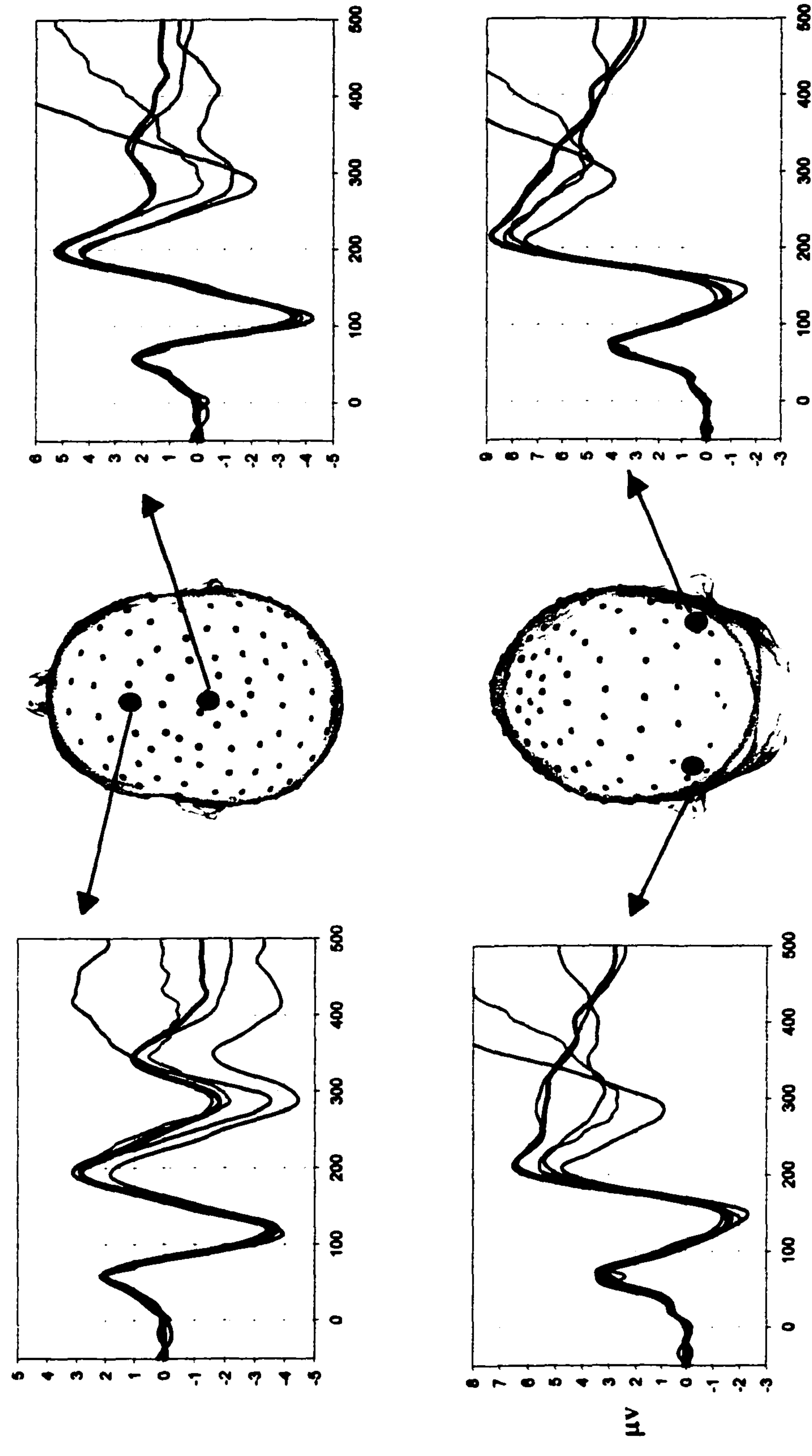
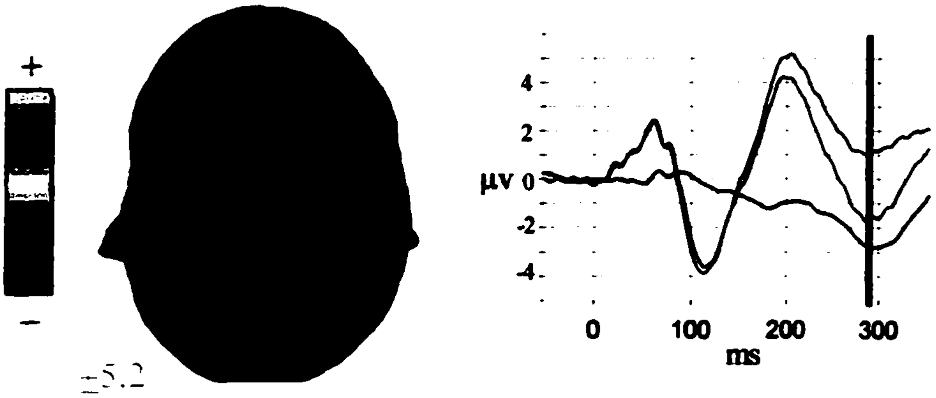

Mean Reaction Time (ms)
Group Cumulative Probability
Miller Inequality
Test 1-
Congruent vs Incongruents
650
600
550
500
450
400
Congr Vis-Incon Aud-Incon
1.0
0.0
0
40
80
Percentile
0.1
0.0
-0.1
0
100
Percentile
Test 2-
Congruent vs Unisensories
650
600
550
500
450
400
Congr Vis-Uni Aud-Uni
1.0
0.0
0
40
80
0.1
0.0
-0.1
0
100

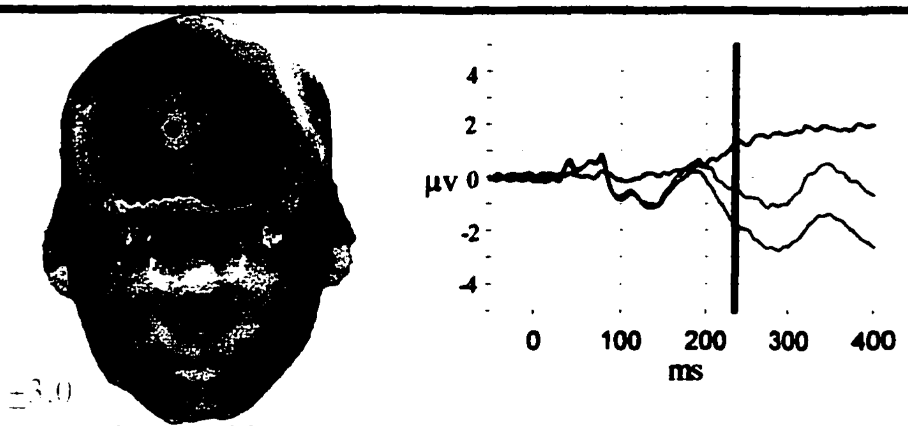
Figure 4. Traces from the 5 bisensory stimuli ERP waveforms are shown for electrodes from fronto-central (upper left), central (upper right), left occipital (lower left), and right occipital (lower right) scalp: these are the congruent-bisensory target ERP (red trace), the visual-incongruent-bisensory target ERP (green trace), the auditory-incongruent-bisensory target ERP (blue trace), the congruent-bisensory non-target ERP (thick black trace), and the incongruent-bisensory non-target ERP (thin black trace).


6 5 4 3 2 1 0 -1 -2 -3 -4 -5
0 100 200 300 400 500
9 8 7 6 5 4 3 2 1 0 -1 -2 -3
0 100 200 300 400 500
5 4 3 2 1 0 -1 -2 -3 -4 -5
0 100 200 300 400 500
8 7 6 5 4 3 2 1 0 -1 -2 -3
0 100 200 300 400 500
ms
µV

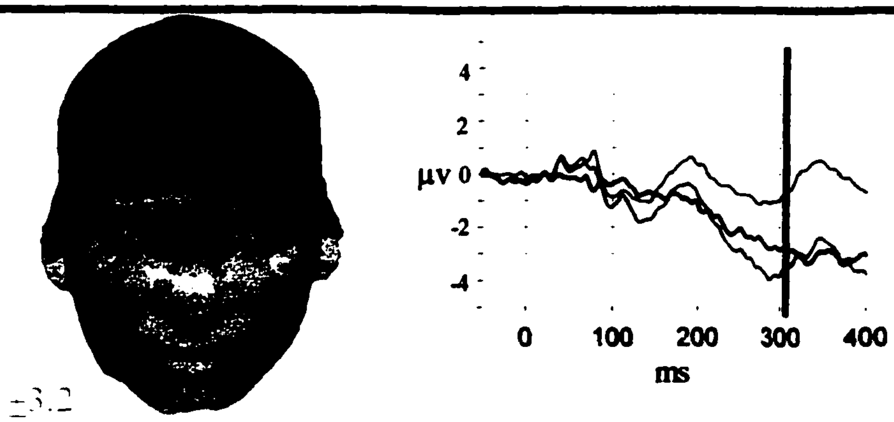
Figures 5-7. Voltage maps and corresponding ERP waveforms are displayed. The black dot indicates the placement of the electrode that the waveforms came from. A straight black line is placed over the waveforms, at the latency the voltage map corresponds to.



Bisensory stimuli: with target auditory element (green trace) versus without target auditory element (red trace). The voltage map (@ 292 ms) and black trace represent the difference.



Bisensory stimuli: with target visual element (green trace) versus without target visual element (red trace). The voltage map (@ 230 ms) and black trace represent the difference.



Bisensory stimuli: with auditory-incongruent target (red trace) versus targets with target visual element (green trace). The voltage map (@ 304 ms) and black trace represent the difference.

CHAPTER 4

Discussion

This thesis examined the cortical spatio-temporal dynamics of visual-auditory multisensory processing. The goal was to first describe neural visual-auditory multisensory interactions under the most simple of stimulus and task conditions. We next asked when and where visual and auditory object information would affect such processes. Data from the first study (Chapter 2) showed a surprisingly early right parieto-occipital AV interaction, which was consistent with the finding of an earlier study (Giard et al., 1999). The timing of onset of this effect (46 ms) was essentially simultaneous with the onset of visual cortical processing, as indexed by the onset of the visual C1 component, which is thought to represent the earliest cortical visual evoked potential. The coincident timing of the early AV interaction and C1 strongly suggests that AV interactions can affect early visual sensory processing. Additional AV interactions were found within the time course of sensory processing (up to 200 ms post stimulus onset). In total, this system of AV effects over the scalp was suggestive of both activity unique to multisensory processing, and the modulation of "unisensory" activity. RTs to the stimuli when presented simultaneously were significantly faster than when they were presented alone. This RT facilitation could not be accounted for by probability summation, as evidenced by violation of the 'race' model, providing compelling evidence that auditory-visual neural interactions give rise to this RT effect.

Visual-auditory interactions over occipito-temporal scalp areas in the latency of the visual N1 in the first study suggested to us that auditory information could affect visual object recognition processes. This finding in conjunction with evidence of direct anatomical connections from auditory to visual areas (Falchier et al., 2001; Rockland &

Ojima, 2001) led us to hypothesize that ancillary auditory information would influence visual object recognition processes, and that this would be reflected in the modulation of the visual N1 and ensuing visual selective attention processes. These predictions were born out in the study on visual-auditory object recognition (Chapter 3). Object recognition was speeded when the visual and auditory elements were object congruent. This RT facilitation could not be accounted for by probability summation, as evidenced by violation of the 'race' model, indicating that visual-auditory interactions gave rise to this RT effect. This was accompanied by the enhancement of cortical processing for the same congruent target stimuli, in visual object recognition areas in the latency range of the visual N1, and over occipital scalp in the latency range of the selection negativity. There were also later occurring visual-auditory congruency effects on the non-target stimuli, which indicated that the non-targets had been processed for meaning.

The findings of this thesis suggest that multisensory integration occurs at many levels of processing, and is affected by task parameters. Further, it shows that there is considerable multisensory processing in sensory specific cortical areas, which occur relatively early on in information processing. While the first study described a spatio-temporal system of visual-auditory multisensory interactions, the second study showed that multisensory interactions are affected by the congruency of featural and object level information between the visual and auditory elements of the stimuli. Future research will be directed at clarifying how the integration of visual and auditory inputs affects the processing of feature and identity information of objects.

CHAPTER 5

Bibliography

Alho, K., Lavikainen, J., J., Reinikainen, K., Sams, M., & Naatanen, R. (1990). Event-related brain potentials in selective listening to frequent and rare stimuli. Psychophysiology, 27, 73-86.

Allison, T., Puce, A., Spencer, D., & McCarthy, G. (1999). Electrophysiological studies of human face perception I: Potentials generated in occipitotemporal cortex by face and non-face stimuli. Cerebral Cortex, 9, 415-430.

Andersen, R.A., Snyder, L.H., Bradley, D.C., & Xing, J. (1997). Multimodal representation of space in the posterior parietal cortex and its use in planning movements. Annual Review of Neuroscience, 20, 303-330.

Anllo-Vento, L., & Hillyard, S.A. (1996). Selective attention to the color and direction of moving stimuli: Electrophysiological correlates of hierarchical feature selection. Perception & Psychophysics, 58, 191-206.

Amedi, A., Malach, R., Hendler, T., Peled, S., & Zohary, E. (2001). Visuo-haptic object-related activation in the ventral visual pathway. Nature Neuroscience, 4, 324-330.

Benevento, L. A., Fallon, J., Davis, B. J., & Rezak, M. (1977). Auditory--visual interaction in single cells in the cortex of the superior temporal sulcus and the orbital frontal cortex of the macaque monkey. Experimental Neurology, 57, 849-872.

Bentin, S., Mouchetant-Rostaing, Y., Giard, M.H., Echallier, J.F., & Pernier, J. (1999). ERP manifestations of processing printed words at different psycholinguistic levels: time course and scalp distribution. Journal of Cognitive Neuroscience 1999 May;11, 235-260

Berman, R.A., & Colby, C.L. (in press). Both auditory and visual attention modulate motion processing in area MT+. Cognitive Brain Research.

Bertelson, P. & Radeau, M. (1981). Cross-modal bias and perceptual fusion with auditory-visual spatial discordance. Perception & Psychophysics, 6, 578-584.

Bruce, C., Desimone, R., & Gross, C.G. (1981). Visual properties of neurons in a polysensory area in superior temporal sulcus of the macaque. Journal of Neurophysiology, 46, 369-384.

Buchner, H., Gobbele, R., Wagner, M., Fuchs, M., Waberski, T.D., & Beckmann, R. (1997). Fast visual evoked potential input into human area V5. Neuroreport, 28, 2419-2422.

Calvert, G.A., Bullmore, E.T., Brammer, M.J., Campbell, R., Williams, S.C., McGuire, P.K., Woodruff, P.W., Iversen, S.D., & David, A.S. (1997). Activation of auditory cortex during silent lipreading. Science, 276, 593-596.

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., Brammer, M.J., & Iversen, S.D. (1998). Crossmodal identification. Trends in Cognitive Sciences, 2, 247-253.

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, 649-57.

Calvert, G.A., Hansen, P.C., Iversen, S.D., & Brammer, M.J. (2001). Detection of audio-visual integration sites in humans by application of electrophysiological criteria to the BOLD effect, 14, 427-438.

Campbell, R., and Dodd, B. (1980) Hearing by eye. Quarterly Journal of Experimental Psychology, 32, 85-99.

Celesia, G.G., & Puletti, F. (1971). Auditory input to the human cortex during states of drowsiness and surgical anesthesia. Electroencephalography and Clinical Neurophysiology, 31, 603-609.

Clark, V.P., Fan, S., & Hillyard, S.A. (1995). Identification of early visual evoked potential generators by retinotopic and topographic analyses. Human Brain Mapping, 2, 170-187.

Clark, V.P., & Hillyard, S.A. (1996). Spatial selective attention affects early extrastriate but not striate components of the visual evoked potential. Journal of Cognitive Neuroscience, 8, 387-402.

Czigler, I. & Balázs, L. (2001). Event-related potentials and audtiovisual stimuli: multimodal interactions. NeuroReport, 12, 223-226.

de Gelder, B., Bocker, K.B., Tuomainen, J., Hensen, M., & Vroomen, J (1999). The combined perception of emotion from voice and face: early interaction revealed by human electric brain responses. Neuroscience Letters, 260,133-136.

Delattre, P.C., Liberman, A.M., & Cooper, F.S. (1955). Acoustic loci and transitional cues for consonants. The Journal of the Acoustical Society of America, 27, 269-273.

Desimone, R., & Gross, C.G. (1979). Visual areas in the temporal cortex of the macaque. Brain Research, 178, 363-380.

Donchin, E., Tueting, P., Ritter, W., Kutas, M., & Heffley, E. (1975). On the independence of the CNV and the P300 components of the human averaged evoked potential. Electroencephalography and Clinical Neurophysiology, 38, 449-461.

Doniger, G.M., Foxe, J.J., Murray, M.M., Higgins, B.A., Snodgrass, J.G., Schroeder, C.E., & Javitt, D.C. (2000). Activation time-course of ventral visual stream object-recognition areas: High density electrical mapping of perceptual closure Processes. Journal of Cognitive Neuroscience, 12, 615-621.

Doniger, GM., Foxe, J.J., Schroeder, C.E., Murray, MM., Higgins, B.A., & Javitt, D.C. (2001). Visual perceptual repetition priming study using high-density electrical mapping. Neuroimage, 13, 305-313.

Driver, J. Spence, C. (2000). Multisensory perception: beyond modularity and Convergence. Current Biology, 10, R731-735.

Duhamel, J.R., Colby, C.L., and Goldberg, M.E. Congruent representation of visual and somatosensory space in single neurons of monkey ventral intraparietal cortex (VIP). In: Brain and Space, (Ed. J. Paillard), Oxford University press, 1991, pp. 223-236.

Eimer, M. (2000). Effects of face inversion on the structural encoding and recognition of faces; Evidence from event-related brain potentials. Cognitive Brain Research, 10, 145-158.

Ernst, M.O., & Banks, M.S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. Nature, 415, 429-433.

Fabiani, M., Kazmerski, V.A., Cycowicz, Y.M., & Friedman, D. (1996). Naming norms for brief environmental sounds: effects of age and dementia. Psychophysiology, 33, 462-475.

Falchier, A., Renaud, L., Barone, P., Kennedy, H. (2001). Extensive projections from the primary auditory cortex and polysensory area stp to peripheral area V1 in the macaque. Society for Neuroscience Abstracts, Vol. 27, Program No. 511.21.

Foxe, J.J., Simpson, G.V., & Ahlfors, S.P. (1998). Cued shifts of intermodal attention: Parieto-occipital ~10 Hz activity reflects anticipatory state of visual attention Mechanisms. Neuroreport, 9, 3929-3933.

Foxe, J.J. Morocz, I..A., Higgins, B.A., Murray, M.M., Javitt, D.C. & Schroeder, C.E. (2000). Multisensory auditory-somatosensory interactions in early cortical processing revealed by high density electrical mapping. Cognitive Brain Research, 10, 77-83.

Foxe, J.J., & Simpson, G.V. (2002). Flow of activation from V1 to frontal cortex in humans: A framework for defining "early" visual processing. Experimental Brain Research, 142, 139-150.

ffytche DH, Guy CN, & Zeki S (1995). The parallel visual motion inputs into areas V1 and V5 of human cerebral cortex. Brain, 118, 1375-1394.

Fu, K.G., Foxe, J.J., Murray, M.M., Higgins, B.A., Javitt, D.C., & Schroeder, C.E. (2001). Cross-modality cued attention-dependent suppression of distracter visual input indexed by anticipatory parieto-occipital alpha-band oscillations. Cognitive Brain Research, 12, 145-152.

Gemba, H. & Sasaki, K. (1989). Potential related to no-go reaction of go/no-go hand movement task with color discrimination in human. Neuroscience Letters, 101, 263-268.

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.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. Perception & Psychophysics, 28, 267-283.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. Perception & Psychophysics, 38, 299-310.

Grunewald, A., Linden, J.F., & Andersen, R.A. (1999). Responses to auditory stimuli in macaque lateral intraparietal area I. Effects of training. Neurophysiology, 82, 330-342.

Guthrie, D., & Buchwald, J.S. (1991). Significance testing of difference potentials. Psychophysiology, 28, 240-244.

Hansen, J.C., Dickstein, P.W., Berka, C., & Hillyard, S.A. (1983). Event-related potentials during selective attention to speech sounds. Biological Psychology, 16, 211-224.

Hansen, J.C., & Hillyard, S.A. (1980). Endogenous brain potentials associated with selective auditory attention. Electroencephalographic and Clinical Neurophysiology, 49, 277-290.

Harter, M.R., & Aine, C. (1984). Brain mechanisms of visual selective attention. On: Parasiraman, P., David, D.R (Eds) : Varieties of Attention. New York: Academic Press, pp. 293-321.

Hikosaka, K., Iwai, E., Saito, H., & Tanaka, K. (1988). Polysensory properties of neurons in the anterior bank of the caudal superior temporal sulcus of the macaque monkey. Journal of Neurophysiology, 60, 1615-1637.

Hillyard, S.A., Hink, R.F., Schwent, V.L., & Picton, T.W. (1973). Evoked potential correlates of auditory signal detection. Science, 182, 177-180.

Hyvarinen, J., & Shelepin, Y (1979). Distribution of visual and somatic functions in the lparietal associative area 7 of the monkey. Brain Research, 169, 561-564.

James, T.W., Humphrey, G.K., Gati, J.S., Servos, P., Menon, R.s., & Goodale, M.A. (in press). Haptic study of three-dimensional objects activates extrastriate visual areas. Neuropsychologia.

Jousmaki, V., & Hari, R. (1998). Parchment-skin illusion: sound-biased touch. Current Biology, 12, R190

Kenemans, J.L., Kok, A., & Smulders, F.T.Y. (1993). Event-related potentials to conjunctions of spatial frequency and orientation as a function of stimulus parameters and response requirements. Electroencephalographic and Clinical Neurophysiology, 88, 51-63.

Kinchla, R. (1974). Detecting target elements in multielement arrays: A confusability model. Perception & Psychophysics, 15, 149-158.

Lewis, J.W., Beauchamp, M.S., & DeYoe, E.A. (2000). A comparison of visual and auditory motion processing in human cerebral cortex. Cerebral Cortex, 10, 873-888.

Macaluso, E., Frith, C.D., & Driver, J. (2000). Modulation of human visual cortex by crossmodal spatial attention. Science, 289, 1206-1208.

MacGrath, M. & Summerfield, Q. (1985). Intermodal timing relations and audio-visual speech recognition by normal-hearing adults. Journal of the Acoustical Society of America, 77, 678-685.

MacLeod, A, & Summerfield, Q. (1990). A procedure for measuring auditory and audio-visual speech-reception thresholds for sentences in noise: rationale, evaluation, and recommendations for use. British Journal of Audiology, 24, 29-43.

Massaro, D. W. (1999). Speechreading: illusion or window into pattern recognition. Trends in Cognitive Sciences, 3, 310-317.

Mazzoni, P, Bracewell, R. M., Barash, S. & Andersen, R. A. (1996). Spatially tuned auditory responses in area LIP of macaques performing delaued memory saccades to acoustic targets. Journal of Neurophysiology, 75, 1233-1241.

McDonald, J.J., & Ward, L.M. (2000). Involuntary listening aids seeing: evidence from human electrophysiology. Psychological Science, 11, 167-171.

McGurk, H. & MacDonald, J. (1976). Hearing lips and seeing voices. Nature, 264. 746-748.

Meredith, M.A., Stein, B.E. (1996) Spatial determinants of multisensory integration in cat superior colliculus neurons. Journal of Neurophysiology, 75, 1843-1857.

Meredith, M.A., Nemitz, J.W., & Stein, B.E. (1987). Determinants of multisensory integration in superior colliculus neurons. I. Temporal Factors. The Journal of Neuroscience, 3215-3229.

Mesulam, M. M. (1981). A cortical network for directed attention and unilateral neglect. Annals of Neurology, 10, 309-325.

Miller, J. (1982). Divided attention: evidence for coactivation with redundant signals. Cognitive Psychology, 14, 247-279.

Miniussi, C., Girelli, M., & Marzi, C.A. (1998). Neural site of the redundant target effect electrophysiological evidence. Journal of Cognitive Neuroscience,10, 216-230.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., & Foxe, J.J. (in press). Multisensory auditory-visual interactions during early sensory processing in humans: A high-density electrical mapping study. Experimental Brain Research.

Movellan, J.R., & McClelland, J.L. (2001). The Morton-Massaro law of information integration: implications for models of perception. Psychological Review, 108, 113-148.

Murray, M. M., Foxe, J. J., Higgins, B. A., Javitt, D. C., & Schroeder, C. E. (2001). Visuo-spatial response interactions in early cortical processing during a simple reaction time task: A high density electrical mapping study. Neuropsychologia, 39, 828-844.

Näätänen, R. (1992). Attention and brain function. Hillsdale, NJ: Lawrence Erlbaum Associates.

Näätänen, R., Gaillard, A.W.K., & Mantysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. Acta. Psychol. (Amst.), 42, 313-329.

Olson, I.R., Gatenby, J.C., & Gore, J.C. (in press). A comparison of bound and unbound audio-visual information processing in the human cerebral cortex. Cognitive Brain Research.

Pardo, J. V., Fox, P.T., & Raichle, M.E. (1991). Localization of a human system for sustained attention by positron emission tomography. Nature, 349, 61-64.

Picton, T.W., Hillyard, S.A., Krausz, H.I., & Galambos, R. (1974). Human auditory evoked potentials. I. Evaluation of components. Electroencephalography and Clinical Neurophysiology, 36, 179-190.

Pourtois, G., de Gelder, B., Vroomen, J., Rossion, B., & Crommelinck, M. (2000). The time-course of intermodal binding between seeing and hearing affective information. Neuroreport, 11, 1329-1333.

Raij, T., Uutela, K., & Hari, R. (2000). Audiovisual integration of letters in the human brain. Neuron, 28, 617-625.

Ritter, W., Simson, R., & Vaughan, H.G.Jr. (1983). Event-related potential correlates of two stages of information processing in physical and semantic discrimination tasks. Psychophysiology, 20, 168-179.

Rockland, K. S., Ojima, H. (2001). Calcarine area V1 as a multimodal convergence area, Society for Neuroscience Abstracts, Vol. 27, Program No. 511.20.

Rossion, B., Gauthier, I., Tarr, M.J., Despland, P., Bruyer, R., Linotte, S., & Crommelinck, M. (2000). The N170 occipito-temporal component is delayed and enhanced to inverted faces but not to inverted objects: an electrophysiological account of face-specific processes in the human brain. NeuroReport, 11, 69-74.

Sasaki, K., & Gemba, H. (1986). Electrical activity in the prefrontal cortex specific to no-go reaction of conditioned hand movement with colour discrimination in the monkey, Experimental Brain Research, 64 603-606.

Sasaki, K., & Gemba, H. (1989). Suppression of visually-initiateed hand movement by stimulation of the prefrontal cortex in the monkey, Brain Research, 495, 100-107.

Schmolesky, M.T., Wang, Y., Hanes, D.P., Thompson, K.G., Leutgeb, S., Schall, J.D., & Leventhal, A.G. (1998). Signal timing across the macaque visual system. Journal of Neurophysiology, 76, 3272-3278.

Schroeder, C.E., & Foxe, J.J. (in press). The timing and laminar profile of converging inputs to multisensory areas of the macaque neocortex. Cognitive Brain Research.

Schroeder, C.E., Lindsley, R.W., Specht, C., Marcovici, A., Smiley, J.F., & Javitt, D.C. (2001). Somatosensory input to auditory association cortex in the macaque monkey. Journal of Neurophysiology, 85, 1322-1327.

Schroeder, C. E., Mehta, A.D., Foxe, J.J. (2001). Determinants and mechanisms of attentional modulation of neural processing. Frontiers in Bioscience, 6, D672-684.

Schroeder, C.E., Mehta, A.D., and Givre, S.J. A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cerebral Cortex, 1998, 8, 575-592.

Selzer, B. & Pandya, D. N. (1980). Converging visual and somatic sensory cortical input to the intraparietal sulcus of the rhesus monkey. Brain Research, 192, 339-351.

Spence, C. , Ranson, J., Driver, J. (2000). Cross-modal selective attention: on the difficulty of ignoring sounds at the locus of visual attention. Perception & Psychophysics, 62, 410-424.

Smid, H.G.O.M., Jakob, A., & Heinze, H.-J. (1999). An event-related brain potential study of visual selective attention to conjunctions of color and shape. Psychophysiology, 36, 264-279.

Snodgrass, J.G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. J Exp Psychology: Human Learning, 6, 174-215.

Stein, B.E., Huneycutt, W.S., & Meredith, M.A. (1988). Neurons and behavior: the same rules of multisensory integration apply. Brain Research, 448, 355-358

Stein, B.E., London, N., Wilkinson, L.K., & Price, D.D. (1996) Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. Journal of Cognitive Neuroscience, 8, 497-506.

Stein, B.E., & Meredith, M.A. (1993). The Merging of the Senses. Cambridge, MA: MIT Press.

Sumby, W.H., & Pollack, I. (1954). Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am., 26, 212-215.

Supek, S., Aine, C.J., Ranken, D., Best, E., Flynn, E.R., & Wood, C.C. (1999). Single vs. paired visual stimulation: superposition of early neuromagnetic responses and retinotopy in extrastriate cortex in humans. Brain Research, 830, 43-55.

Talsma, D., & Kok, A. Nonspatial intermodal selective attention is mediated by sensory brain areas: Evidence from event-related potentials. Psychophysiology, 38, 736-751.

Teder-Sälejärvi, W. A., McDonald, J. J., Di Russo, F., & Hillyard, S. A. (in press). An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings. Cognitive Brain Research.

Thompson, L.A. (1995) Encoding and memory for visible speech and gestures: a comparison between young and older adults. Psychology and Aging, 10(2): 215-228.

Vaughan, H.G., & Arezzo, J.C. (1988). The neural basis of event-related potentials. In T.W. Picton (Ed.), Human event-related potentials. Handbook of Electroencephalography and Clinical Neurophysiology. Revised series, (Vol. 3, pp. 45-96). Amsterdam: Elsevier.

Vaughan, H.G., & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. Electroencephalography and Clinical Neurophysiology, 28, 360-367.

Vaughan, H.G., Ritter, W., & Simson, R. (1980). Topographic analysis of auditory event-related potentials. Progress in Brain Research, 54, 279-285.

Vincent, S.B. (1912). The function of the vibrissae in the behavior of the white rat, Behavioral Monographs, 1.

Vogel, E.K., Luck, S.J. (2000). The visual N1 component as an index of a discrimination process. Psychophysiology,37, 190-203.

Wallace, M.T., Meredith, M.A., & Stein, B.E. (1992). Integration of multiple sensory modalities in cat cortex. Experimental Brain Research, 91, 484-488.

Walter, W.G., Cooper, R., Alderidge, V.J., McCallum, W.C., & Winter, A.L. (1964). Contingent negative variation: An electric sign of sensorimotor association and expectancy in the human brain. Nature, 203, 380-384.

Woods, D.L., & Alain, C. (2001). Conjoining three auditory features: an event-related brain potential study. Journal of Cognitive Neuroscience, 15, 492-509

Woods, D.L., Alho, K., & Algazi, A. (1992). Intermodal selective attention. I: Effects on event-related potentials to lateralized auditory and visual stimuli. Electroencephalography and Clinical Neurophysiology, 82, 341-355.

Woods, D.L., Hillyard, S.A., & Hansen, J.C. (1984). Event-related potentials reveal similar attentional mechanisms during selective listening and shadowing. Journal of Experimental Psychology: Human Perception & Performance, 10, 761-777.