Congruency of auditory sounds and visual letters modulates mismatch negativity and P300 event-related potentials

Allison J.D. Andres [a], Janis E. Oram Cardy [a], Marc F. Joanisse [b],*

[a] School of Communication Sciences and Disorders, The University of Western Ontario, London, Ontario, Canada
[b] Department of Psychology and Centre for Brain and Mind, The University of Western Ontario, London, Ontario, Canada

ABSTRACT

A key determinant of skilled reading is the ability to integrate the orthographic and auditory forms of language. A number of prior studies have identified neural markers in adult readers corresponding to audio–visual integration of letters and their corresponding sounds. However, there remains some controversy as to the stage of processing at which this occurs. In the present study, we examined this issue using event-related potentials (ERPs), due to their sensitivity to the timing of perceptual and cognitive processes. Letter sounds were presented auditorily in an unattended mismatch negativity (MMN) paradigm, which is argued to be indicative of auditory sensory memory. Concurrently, participants performed a visual letter identification task. On critical trials, the auditory stimulus was played concurrently with the visual letters. We observed significant MMNs both when the visual letter was congruent with the auditory stimulus, and when it was incongruent. However, the magnitude and scalp distribution of this effect was attenuated in incongruent trials. We also observed a later-going effect of congruency on P300 trials, marked by increased amplitudes and latencies for incongruent compared to congruent trials. The results suggest audiovisual integration of letters and sounds can and does occur during relatively early pre-attentive stages of sensory processing, and that these effects extend to later-going attentional phases of processing as well.

© 2010 Elsevier B.V. All rights reserved.

1. Introduction

A key element of human perception is the ability to integrate concurrent sensory information across different modalities. Of particular interest in language research is audiovisual (AV) integration. This ability plays an important part in reading, the development of which hinges on the ability to tie together visual letters with auditory phonemes. A failure to develop these connections has been proposed as a factor in dyslexia, and training programs involving the presentation of phonemes and visual letters have been found to improve reading ability in children with developmental dyslexia (Magnan et al., 2004).

Similarly, integration of visual cues, such as lip movements, with auditory information can change auditory speech perception. For example, Sumby and Pollack (1954) found that speech perception in noise was significantly improved when participants were able to see the corresponding lip movements. Likewise, visual information alters speech perception in the McGurk illusion, in which participants are presented with an auditory stimulus (e.g., /ga/) paired with a face making incongruous lip movements (e.g., /ba/). In this paradigm,

listeners report perceiving the auditory sound incorrectly (e.g., hearing /ba/ or /da/; McGurk and MacDonald, 1976).

Recent research has focused on AV integration in speech perception as it pertains to the point in time during perception at which cross-modal stimuli are merged onto a single percept. Traditionally, visual and auditory information have been thought to be processed independently in their respective sensory cortices prior to being combined within specific multisensory regions. From this perspective, AV presentation of stimuli impacts primary sensory cortices due to feedback from multisensory sites (e.g., van Atteveldt et al., 2004). In contrast, other research suggests that simultaneously presenting auditory and visual information can affect the early sensory processing of stimuli in their respective sensory cortices, possibly through feed-forward inputs and lateral connections between primary sensory areas (Molholm et al., 2002; see Foxe and Schroeder, 2005 for review).

Currently, evidence of early vs. late AV integration in language stimuli is equivocal. Raij et al. (2000) examined the integration of visually presented letters and auditorily presented phonemes using magnetoencephalography (MEG). Matching (congruent) and non-matching (incongruent) letters and sounds were presented to participants. These concurrent audiovisual conditions were compared to the sum of unimodally presented auditory and visual letter stimuli, and any differences in cortical activation between the audiovisual conditions and the summed unimodal conditions were considered

* Corresponding author. Department of Psychology, The University of Western Ontario, London Ontario Canada, N6A 5C2. Tel.: +1 519 661 2111x86582.
E-mail address: marcj@uwo.ca (M.F. Joanisse).

interactions. They found no AV interactions in sensory-specific cortices such as primary auditory cortex. However, presentation of congruent auditory and visual stimuli yielded superadditive effects localized to the left occipito-temporal region as well as in left and right superior temporal sulci (STS); likewise, these interactions occurred at a time window later than would be expected for primary sensory processing (i.e., at 465 ± 50 ms post stimulus and 495 ± 50 ms post stimulus, respectively). Such findings seem in line with the view of integration occurring at a higher-order 'multisensory' level during processing.

Conversely, many studies support the view that simultaneously presented auditory and visual information can affect early sensory processing (Foxe and Schroeder, 2005; Widmann et al., 2004; Teder-Sälejärvi et al., 2002). Most studies of this type have focused on simple auditory and visual stimuli. However, similar effects have also occurred for letter-sound stimulus pairs as well; Herdman et al. (2006) used MEG to examine congruent and incongruent presentations of letters and phonemes. They found larger response power in left auditory cortex on congruent trials, relative to incongruent trials, early in perceptual processing of concurrent audio and visual stimuli (0 to 250 ms post stimulus). These results indicate that concurrent auditory and visual information affects early auditory processing, supporting an early integration model.

The current study examines whether integration of visual letters and letter sounds occurs in a time period consistent with early, pre-attentive integration, or later in processing, using a design that has several key differences from previous studies in this area. One concern common to earlier MEG studies of AV integration is the use of an overt monitoring task. Prior studies have asked listeners to directly monitor for AV congruency, which directs their attention to both the visual and auditory domains and also emphasizes congruent trials over incongruent trials (Raij et al., 2000; Herdman et al., 2006). It is not clear whether this could have influenced either the nature or timing of brain responses to letter-sound congruency. For instance, Van Atteveldt et al. (2007) found that fMRI activation in temporal lobe regions associated with unattended AV integration of letters and sounds was eliminated when subjects were instructed to actively match visual letters to auditory sounds. In the present study, subjects monitored only the visual letter stimuli, and were instructed to ignore the auditory stream. This allowed us to examine whether congruency effects emerge in ERP results even when listeners are not actively attending to the congruency of the visual and auditory stimuli.

Subjects were instructed to monitor for an experimentally irrelevant feature within the visual stream, namely whether the presented letter was a vowel. A go/no-go task was used in which they pressed a response key only if the letter was a vowel. Notably, all critical trials of interest involved visually presented vowels. This had the key benefit that all of the trials included in our analyses involved generating the same manual response. Comparisons of ERPs across conditions should thus subtract away influences of response selection and execution mechanisms unrelated to the experimental manipulations of interest.

To determine whether the integration of letters and sounds occurs during pre-attentive sensory processing, the mismatch negativity (MMN) ERP component was examined. Our approach builds on findings that the MMN can be modulated by concurrent visual stimuli, even though this component occurs in response to unattended auditory stimuli (Froyen et al., 2008; Sams et al., 1991; Saint-Amour et al., 2007). In an MMN paradigm, a single sound is played repeatedly (*standards*), occasionally interspersed with an oddball stimulus (a *deviant*). The MMN is marked by a negative-going deflection in electrical brain signals in response to the deviant, relative to the repeated standard (Näätänen, 2000; Näätänen et al., 2007). It is suggested that perceiving the standard stimulus leads to the development of a trace in auditory short-term memory. The deviation from this memory trace results in an MMN response in auditory

cortex, which is observed most strongly across the midline frontal and central electrodes. However, Froyen, et al. (2008) found that the magnitude of the MMN was enhanced when the auditory deviant was accompanied by a visual letter that was congruent with the repeated standard (seeing the letter 'a' when hearing the Dutch vowel /a/ as the standard and the Dutch vowel /o/ as the deviant), compared to when no letter was displayed during the standard and deviant trials. This effect was greatest when the auditory and visual streams were simultaneous, but was significantly weaker when the two were temporally asynchronous by 100 or 200 ms.

The Froyen et al. (2008) results suggest that AV congruency can influence automatic and relatively early-going neural responses to auditory speech, favoring the early integration model of AV integration. A key benefit to using an MMN paradigm for the auditory component of this study is that it is an early-going ERP component that can be observed even for unattended stimuli; it is thus generally accepted to reflect early stages of auditory perception occurring prior to conscious decision processes (Näätänen, 2000; Picton et al., 2000). Consequently, Froyen et al. suggest that changes in the MMN in response to an incongruent visual letter indicate that AV integration alters processing at an early and pre-attentive level. Likewise, this effect is not necessarily due to having directed listeners to monitor for an auditory mismatch.

This earlier study, however, only examined the modulation of MMNs in response to the presence of an incongruent visual letter compared to no letter at all. Therefore, MMN modulations could have been due simply to the presence of a concurrent letter stimulus, regardless of its relationship to the standard or deviant sound. In addition, the visual letter was presented repeatedly at the same inter-trial interval as the auditory stream, and the letter stimulus was constant across all trials. As such, subjects were potentially aware of the AV congruency of letters and sounds, which might explain the modulation in the MMN amplitude.

The present study used a similar approach of using MMN to examining AV congruency. However, the visual task paradigm was appreciably different, and permitted us to manipulate both AV congruency and auditory mismatch. Subjects passively heard the auditory MMN stream as they actively performed a visual detection task on rapidly presented letters. The rate of presentation of visual letters differed from that of the auditory mismatch paradigm, which meant that on most trials visual and auditory stimuli were not being heard concurrently. However, the timing was such that auditory and visual stimuli were occasionally presented simultaneously. Moreover, the auditory letter name was either congruent with the visual letter (e.g., E and /i/), or incongruent (e.g., A and /i/). Likewise, the auditory stimulus played concurrently with a letter was either a standard or deviant. This crossing of congruency and mismatch provided us with information about the nature and timing of AV integration of letters and sounds in two ways. First, it allowed us to examine the influence that congruency has on ERP correlates of visual attention during a letter detection task; and second, it allowed us to examine the extent to which congruency also influences the MMN component, which is relatively earlier-going, pre-attentive and automatic.

The behavioral paradigm used here had several advantages over prior studies of AV congruency and MMN. The first was that subjects were not instructed to actively attend to the auditory stream. This was further assured by the relatively high task demands imposed by the speeded visual letter detection task, and by the fact that the auditory stimuli played asynchronously with the visual detection task except in a small number of trials. We also note that the number of standard and deviant trials was better balanced than is typical in MMN studies, due to the fact that only trials involving concurrent letters and sound were analyzed; as a result, standards and deviants had roughly comparable signal-to-noise ratios, something that is problematic in MMN paradigms (Picton et al., 2000).

Of special interest was whether MMN responses would be modulated by concurrent presentation of congruent vs. incongruent visual letters. Variations in MMN response due to changes in letter congruency would indicate that visual information impacts auditory processing pre-attentively, which is in line with research that suggests that AV integration occurs at a very early stage of processing in the auditory cortex. Conversely, if changes in the congruence of the visual information impacted only later parts of the ERP signal, this would be more consistent with research suggesting AV integration following independent sensory processing.

A related question was whether AV congruency would impact ERP components associated with the attended visual detection task, despite instructions to participants to ignore the auditory information. It is known that such congruency effects influence behavioral measures such as reaction time; for example, subjects show faster reaction times to pictures of objects when presented with auditory sounds that match the objects, compared to when sounds are incongruent with pictures (Laurienti et al., 2003). Of interest here is how unattended letter-sound congruency might also influence the P300, a commonly used ERP index of visual attention (for a recent review, see Polich, 2007). A key characteristic of the P300 is its sensitivity to target discrimination; in the present study we were interested in modulations in P300 when subjects are confronted with identifying a visual letter stimulus in the context of unattended congruent vs. incongruent sounds.

2. Method

2.1. Subjects

All procedures were approved by the University of Western Ontario Department of Psychology Research Ethics Board. We recruited 22 neurologically healthy adults (11 female) aged 18 to 35. All were right-handed by self-report, spoke fluent English, and had normal or corrected to normal vision. None had significant hearing loss as assessed using a free-field 45 dB pure tone detection task at 500, 1000, 2000 and 4000 Hz. All subjects provided informed consent prior to participating.

2.2. Stimuli and procedures

Auditory stimuli consisted of the vowels /i/ and /o/ (corresponding to the English names of the vowels E and O) synthesized into 200 ms waveforms, 22,050 Hz sample rate, and 16-bit quantization, using the *Synth* Klatt Parameter Editor (London, ON: Avaaz Innovations). This pair represented the vowel names in English that differed maximally in formant frequencies, and are thus the most likely to yield mismatch effects (Picton et al., 2000). Items were post-processed with a 20 ms linear ramp from zero to full amplitude at onset and a 25 ms linear ramp to zero intensity at offset. To approximate a natural voice contour, fundamental and formant frequencies were varied throughout the stimulus duration (Table 1). Visual stimuli consisted of the

uppercase consonant and vowel letters (A E I O U and R S T L N, respectively) presented in black 24 pt Courier New typeface at the centre of a CRT display.

Testing was performed in a sound treated room, with subjects seated comfortably in front of a CRT display. Subjects performed a vowel identification task in which they were presented with individual visual letters presented for 200 ms, followed by a 600 ms fixation cross. Letter stimuli were presented in pseudorandom order. They were instructed to look at each letter and press the "V" key if the letter was a vowel, and do nothing if it was not a vowel. Prior to performing experimental trials, subjects performed 100 practice trials (50 vowel and 50 consonant trials, presented in random order) to acclimate them to the procedure. All subjects were able to produce an accurate response within the 800 ms trial period at a threshold of 80% or higher on these practice trials.

On experimental trials subjects performed the vowel detection task described above concurrently with hearing an unattended auditory MMN stream (Fig. 1). Testing was divided across two blocks of 1122 visual detection trials each. Auditory items were presented via loudspeakers at 50 dB SPL (measured free-field at the location where the subjects were seated). Items consisted of a repeated standard (772 repetitions per block) played at 1000 ms SOA, pseudorandomly interspersed with an occasional deviant (152 presentations per block). This yielded a rate of approximately 5:1 standards to deviants, with a minimum of three standards separating each deviant. The standard and deviant stimuli were counterbalanced across blocks such that /i/ was the standard and /o/ was the deviant in one block, and the two were reversed in the other block. The order of blocks was randomized across subjects.

Subjects were instructed to ignore the auditory stimuli and focus on responding to the visual letters as quickly and accurately as possible. Importantly, the timing of auditory stimuli and visual detection trials differed such that auditory and visual stimuli were only occasionally presented simultaneously (every 5–6 trials; Fig. 1). Auditory stimuli were presented at 1000 ms SOA, whereas visual trials were presented at 800 ms SOA (reduced to 600 ms SOA every 18 trials; see below). The use of different SOAs minimized the likelihood that subjects would attend to the auditory stream, or to anticipate an auditory stimulus in order to perform the visual task more rapidly or accurately.

'Go' trials in the vowel detection task coincided with both standards and deviants in the auditory stream; half the coincident trials were letter–sound congruent (e.g., seeing E and hearing /i/) and half were incongruent (e.g., seeing A and hearing /i/). The result was four AV trials of interest: congruent standards (156 trials); incongruent standards (162 trials); congruent deviants (100 trials); and incongruent deviants (100 trials). Note also that all four trial types involved a button press response, such that this factor was equated across conditions.

Every 18th visual trial, the SOA was reduced from 800 to 600 ms. This was done to ensure that auditory and visual stimuli coincided appropriately to create the necessary AV pairings.

Table 1
Bandwidth and frequency properties of synthesized Vowels /i/ and /o/.

| Vowel	Bandwidth (Hz)	Formants
/o/	B1 = 80	F0 = 100 Hz at onset, increased linearly to 140 Hz at 30 ms, decreased linearly to 80 Hz at 135 ms, maintained until offset.
	B2 = 70	F1: 560 Hz at onset, maintained until 80 ms, decreased linearly to 440 Hz at offset.
	B3 = 70	F2: 900 Hz at onset, maintained until 60 ms, decreased linearly to 800 Hz at offset.
		F3: 2260 Hz from onset to offset.
		F4: 3300 Hz from onset to offset.
/i/	B1 = 50	F0 = 100 Hz at onset, increased linearly to 140 Hz at 35 ms, decreased linearly to 80 Hz at offset.
	B2 = 200	F1 = 320 Hz at onset, decreased linearly to 220 Hz at 75 ms, maintained until offset.
	B3 = 400	F2 = 2200 Hz at onset, decreased linearly to 2000 Hz at 75 ms, maintained until offset.
		F3 = 3120 Hz at onset, decreased linearly to 2860 Hz at 80 ms, then increased linearly to 2900 Hz at offset.
		F4 = 3300 Hz from onset to offset.
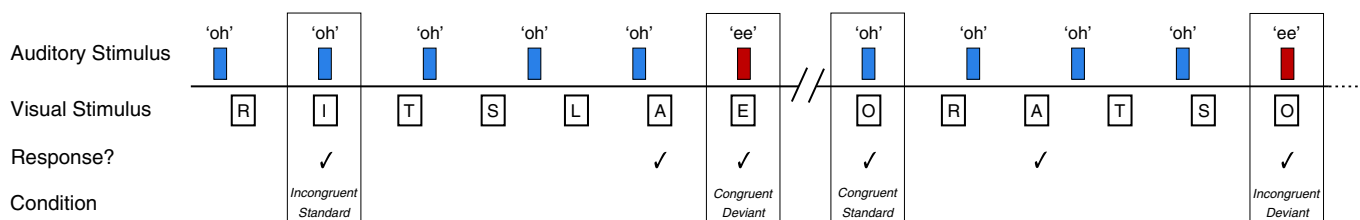
Fig. 1. Schematic diagram of trial presentation paradigm for audio–visual mismatch study. Subjects performed a visual detection task with button press concurrently with a passive auditory mismatch negativity (MMN) paradigm. Presentation rates of the auditory stimuli varied such that they were only presented concurrently with visual letters every 5–6 trials. Critical trials are outlined, and reflect the crossing of audio–visual congruency, and standard/deviant MMN stimuli.

2.3. EEG recording

EEG signals were recorded from 32 channels, digitized at 500 Hz using a Synamps amplifier, filtered online at.1 to 100 Hz with a 60 Hz notch filter, referenced to nose tip. Recordings were obtained using Ag/AgCl sintered electrodes with impedances kept below 10 kΩ; 30 scalp sites were recorded using an electrode placement cap configured in the international 10–20 system. Eye electrodes were placed on the outer canthi and above and below the left eye to record horizontal and vertical eye movements, respectively.

ERP trials were obtained by dividing recordings into epochs -200 to 600 ms post stimulus onset for each of the four AV concurrent trial types (standard congruent, deviant congruent, standard incongruent, and deviant incongruent). Signals were filtered offline with a bandpass zero phase shift digital filter (0.1 to 30 Hz, 12 dB/octave) and baseline corrected to the average voltage of the 200 ms pre-stimulus interval. Trials were rejected for artifacts on the basis of voltages exceeding ± 100 μV at scalp electrodes and ± 75 μV at eye electrodes.

MMN effects were quantified for statistical analyses by calculating the mean voltage from 110 to 190 ms post stimulus onset for each trial type. Nine electrode sites were selected for analyses (Fz, F7, F8, Cz, T7, T8, Pz, P7, and P8), providing coverage of anterior, central, and posterior scalp sites over both hemispheres and at midline. A three-way ANOVA was then used to examine the effects of auditory Mismatch (standard and deviant) and two electrode position factors (Laterality: left, midline and right; and Anterior/posterior: frontal, central/temporal and parietal) with mean voltage as the dependent variable. Greenhouse-Geisser conservative degrees of freedom were used for all ERP analyses, to protect from sphericity violations that are common in multi-electrode data. For clarity, we report all significant main effects but only interactions involving the Mismatch factor. Significant interactions of mismatch and electrode site factors were followed up on with planned comparisons (pairwise t-tests, two-tailed) that compared mean voltages for standard and deviant trials at each electrode site.

Next, the magnitude and latency of MMNs were computed by subtracting standard from deviant waveforms. The amplitude and latency of the peak of each subtraction waveform were calculated subjectwise using the point of greatest negativity within the 110–190 ms time interval, separately for each electrode of interest. The peak voltages and peak latencies were submitted to a three-way repeated measures ANOVA for the effects of Congruency (congruent and incongruent) and the two electrode factors (Anterior/posterior and Laterality).

P300 effects were analyzed by calculating the amplitude and latency of the most positive-going peak within the 360–600 ms post stimulus onset. We compared congruency effects separately for the standard and deviant trials, using three-way ANOVAs for the effects of Congruency, Laterality and Anterior/posterior. We focused these analyses on congruency based on the expectation that this factor would be the most likely to exert an influence on the P300, given established influences of attention and stimulus expectancy on this component (e.g., Kutas et al., 1977). In the interest of brevity, here

again we report all significant main effects but only those interactions involving the Congruency manipulation. Planned comparisons consisted of pairwise t-tests (two-tailed) that compared the peak amplitude or latency of the congruent and incongruent conditions at each electrode site.

3. Results

3.1. Behavioral data

Detection accuracy and RTs were calculated for the four critical trial types (Table 2), and compared using 2-way repeated measures ANOVAs for the effects of Congruency and Mismatch (standard and deviant). Accuracy data showed a main effect of Mismatch, $F(1,21) = 5.62$, $p < .05$, $\eta2 = 0.211$; there was no significant effect of Congruency and no interaction. Analysis of RT data showed main effects of Congruency and Mismatch, $F(1,21) = 11.27$, $p < .01$, $\eta2 = 0.349$, and $F(1,21) = 75.37$, $p < .001$, $\eta2 = 0.782$, qualified by a significant interaction, $F(1,21) = 35.20$, $p < .001$, $\eta2 = 0.626$. Follow-up analysis of the interaction using paired samples t-tests (two-tailed), revealed faster RTs for Deviant Congruent trials compared to both Standard Congruent and Deviant Incongruent trials ($Ps < .001$), and for Standard Congruent vs. Standard Incongruent trials ($P < .05$).

3.2. Mismatch negativity

We next examined auditory mismatch effects within the congruent and incongruent conditions (Fig. 2a,b). As is apparent in the figure, a mismatch effect appears to have been evoked during both the congruent and incongruent trials, marked by increased early-going negativity concentrated at the frontal–central electrodes. This was examined by comparing standard and deviant waveforms separately for the congruent and incongruent conditions using a three-way ANOVA for the effect of auditory Mismatch, Anterior/posterior and Laterality (see Method section for details).
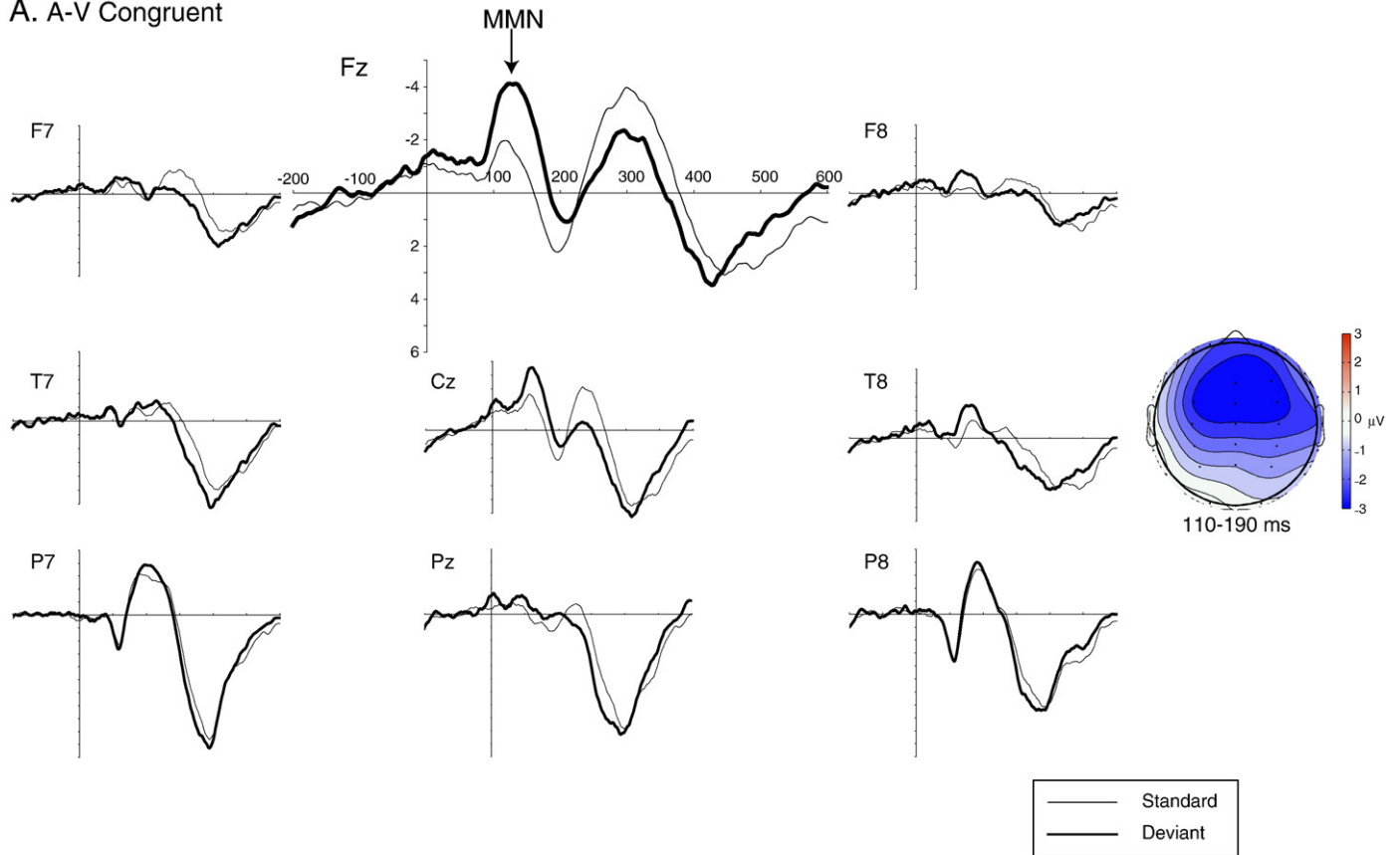
For congruent trials, we observed a significant main effect of Mismatch, $F(1,21) = 18.94$, $p < .001$, $\eta2 = .474$, qualified by significant interactions of Mismatch × Anterior/posterior, $F(2,42) = 13.58$, $p < .001$, $\eta2 = .444$, and Mismatch × Laterality, $F(2,42) = 29.31$, $p < .001$, $\eta2 = .720$. The three-way interaction was also significant, $F(4,84) = 4.74$, $p < .01$, $\eta2 = .492$. Planned comparisons revealed significant differences between standards and deviants at electrodes Fz, Cz, T8 ($Ps < .001$), Pz, and F7 ($Ps < .05$).

Table 2
Behavioral results from the visual detection task. Mean (SE) detection accuracy and reaction times (RTs) are listed for 'Go' responses on vowel trials.

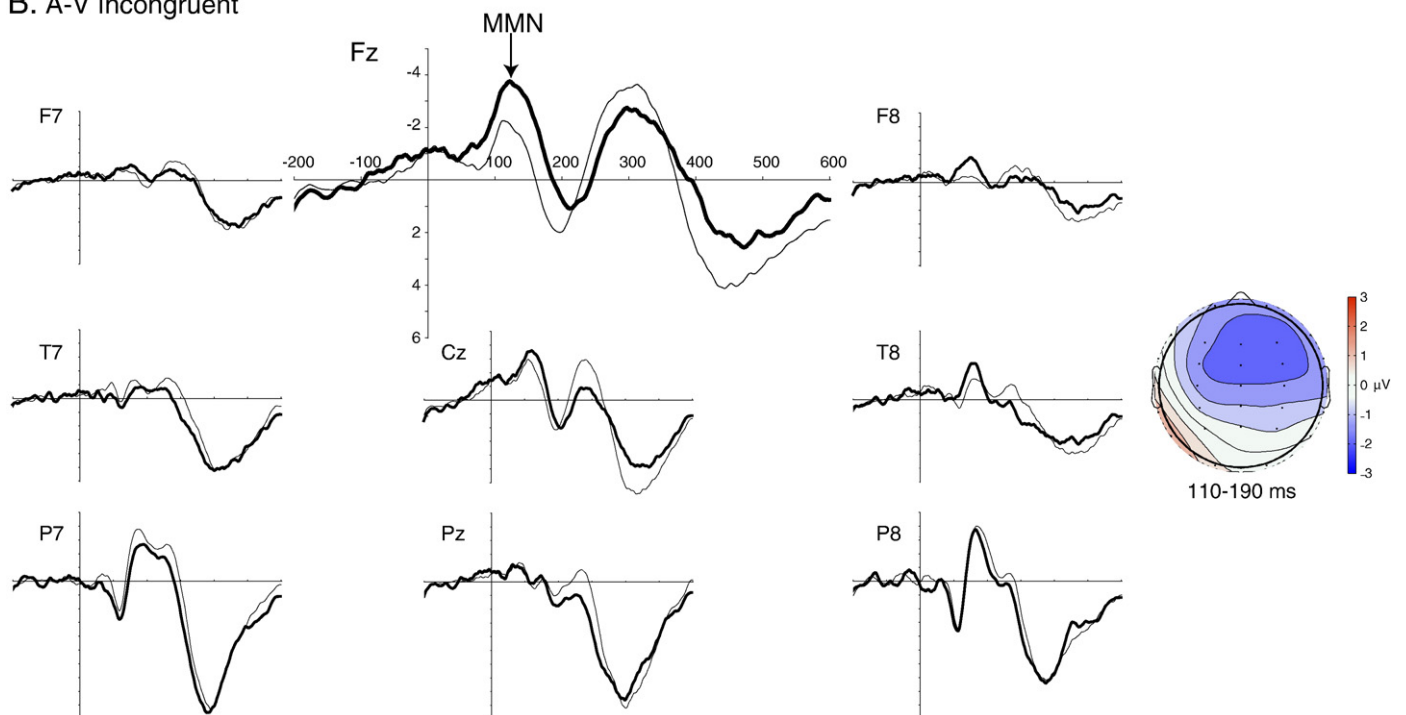| Mismatch	A/V congruency	Proportion correct	RT (ms)
Standard	Congruent	0.95 (.01)	381 (9)
	Incongruent	0.96 (.01)	387 (8)
Deviant	Congruent	0.97 (.01)	361 (9)
	Incongruent	0.97 (.01)	389 (10)

Fig. 2. Comparison of standard and deviant trials. Scalp maps illustrate the electrode distribution of the MMN effect, obtained by subtracting standard from deviant trials. (A) Congruent trials in which visual letter matched the auditory stimulus. (B) Incongruent trials in which the visual letter differed from the auditory stimulus.

Incongruent trials did not show a main effect of Mismatch, however they did reveal significant interactions of Mismatch×Anterior/posterior, $F(2,42) = 26.87$, $\eta2 = .561$, $p < .001$, and Mismatch×Laterality, $F(2,42) = 15.56$, $p < .001$, $\eta2 = .426$. The three-way interac-

tion was also significant, $F(4,84) = 4.97$, $p < .01$, $\eta2 = .191$. Planned comparisons revealed significant differences between standard and deviant at electrodes Fz ($P < .001$), Cz, F8 and P7 ($Ps < .05$); note however that the difference at P7 was in fact in the opposite direction

than would be predicted for an MMN, marked by more positive voltages for the deviant condition.

We followed up on this by comparing the magnitude and latency of MMN effects in the congruent vs. incongruent conditions. MMN effects were calculated as subtraction waves (standard minus deviant; Fig. 3). These were submitted to three-way repeated measures ANOVAs for the effects of Congruency, Anterior/posterior and Laterality. The analysis of peak voltage revealed significant effects for both electrode position factors (Anterior/posterior, $F(2,42) = 11.46$, $p < .01$, $\eta 2 = .578$, and Laterality, $F(2,42) = 31.86$, $p < .001$, $\eta 2 = .591$). There was no main effect of Congruency, however this was qualified by a significant Congruency × Anterior/posterior interaction, $F(2,42) = 5.00$, $p < .05$, $\eta 2 = .190$. Planned comparisons (paired samples t-tests) compared the amplitude of mismatch effects at each electrode site, and revealed significantly greater negativity at Pz, $t(21) = 2.64$, $p < .05$. The difference at electrode P7 just missed significance, $t(21) = 2.02$, $p = .06$. The analysis of MMN latency revealed only a significant main effect of Laterality, $F(2,42) = 3.53$, $p = .05$. There was no significant main effect of congruency or Front/back, and no interactions.

3.3. P300

We next examined how auditory–visual congruency affected the later-going P300 component. This was done separately for the auditory match and mismatch trials, on the assumption that congruency would exert the strongest influence on the P300. As is apparent in Fig. 4, congruency appeared to be influencing the amplitude of the P300 component for auditory standards (Fig. 4a) and latency in the case of auditory deviants (Fig. 4b). This was examined by calculating the amplitude and latency of peak voltages from 360–600 ms post stimulus onset for each trial type. Three-way ANOVAs then examined the effects of Congruency, Laterality and Anterior/posterior on either latency or amplitude.

For standard trials, the analysis of amplitude did not reveal a main effect of Congruency. However, there was a significant Congruency × Laterality interaction, $F(2,42) = 4.34$, $p < .05$, $\eta 2 = .171$; and the three-way interaction also just missed significance, $F(4,84) = 2.69$, $p = .052$, $\eta 2 = .114$. We also observed significant main effects of Laterality and Anterior/posterior, $F(2,42) = 12.18$, $p < .001$, $\eta 2 = .367$, $F(2,42) = 103.47$, $p < .001$, $\eta 2 = .831$, respectively. Interactions of the Congruency and electrode site factors appear to reflect increased P300 amplitudes for the incongruent trials, especially at the frontal and central electrodes along the midline. Consistent with this, planned comparisons revealed significantly greater negativity for incongruent

trials at the Cz electrode site ($P < .05$). The analysis of latency did not reveal main effects or interactions involving Congruency.

Results of the deviant trials yielded somewhat different effects. Analysis of amplitudes did not reveal a main effect or interactions involving Congruency. Instead, we found differences in the latency of the peak, marked by a significant main effect of Congruency, $F(1,21) = 24.40$, $p < .001$, $\eta 2 = .537$, qualified by a Congruency × Anterior/posterior interaction, $F(2,42) = 4.52$, $p < .05$, $\eta 2 = .117$. There were also main effects of Laterality, $F(2,42) = 5.98$, $p < .01$, $\eta 2 = .221$ and Anterior/posterior, $F(2,42) = 57.92$, $p < .001$, $\eta 2 = .734$. Planned comparisons confirmed significantly later peaks for the incongruent condition at electrodes Fz, T7, F7 ($Ps < .001$), Cz ($P < .01$) and Pz ($P < .05$).

4. Discussion

The present study examined the time course of integrating auditory speech and visual orthographic stimuli. Of primary interest was identifying the point at which letter-sound congruency influences perceptual processing. Our paradigm focused on the MMN, an automatic early-going component indexing auditory memory and discrimination, and the P300, a later-going index of visual attention. The MMN paradigm involved presenting a stream of repeated vowels (standards) occasionally interrupted by an oddball vowel (deviant). Subjects were told to ignore the auditory stream and instead to perform a visual detection task in which they viewed letters presented one at a time on a computer display, pressing a button only when a vowel was presented.

Of interest were trials in which target vowel stimuli were presented concurrently with an auditory standard or deviant. Manipulating the congruency of the letter and sound yielded a crossing of the two variables of interest: letter–sound congruency and auditory mismatch. With respect to this second manipulation, we observed significant MMN effects in both the congruent and incongruent conditions, marked by stronger negativity at the fronto-central midline electrodes peaking approximately 150–180 ms post stimulus onset. However, we observed a larger effect in the congruent condition, marked by negativities across a more widespread range of scalp sites. A direct comparison of subtraction waves confirmed that the mismatch effect extended to more posterior electrodes in the congruent condition. The results indicate that the congruency of AV stimuli can modulate the MMN response, such that integration of phonetic and orthographic information can and does occur at a relatively early, pre-attentive stage in stimulus processing.
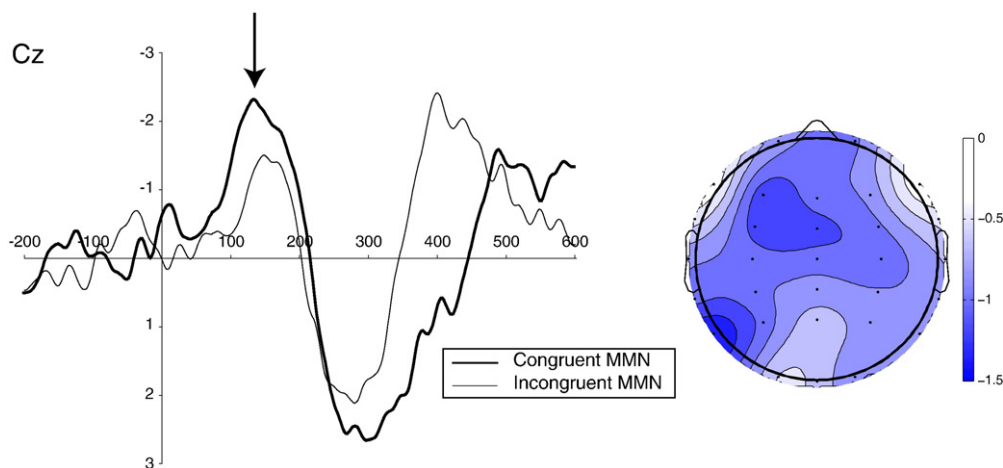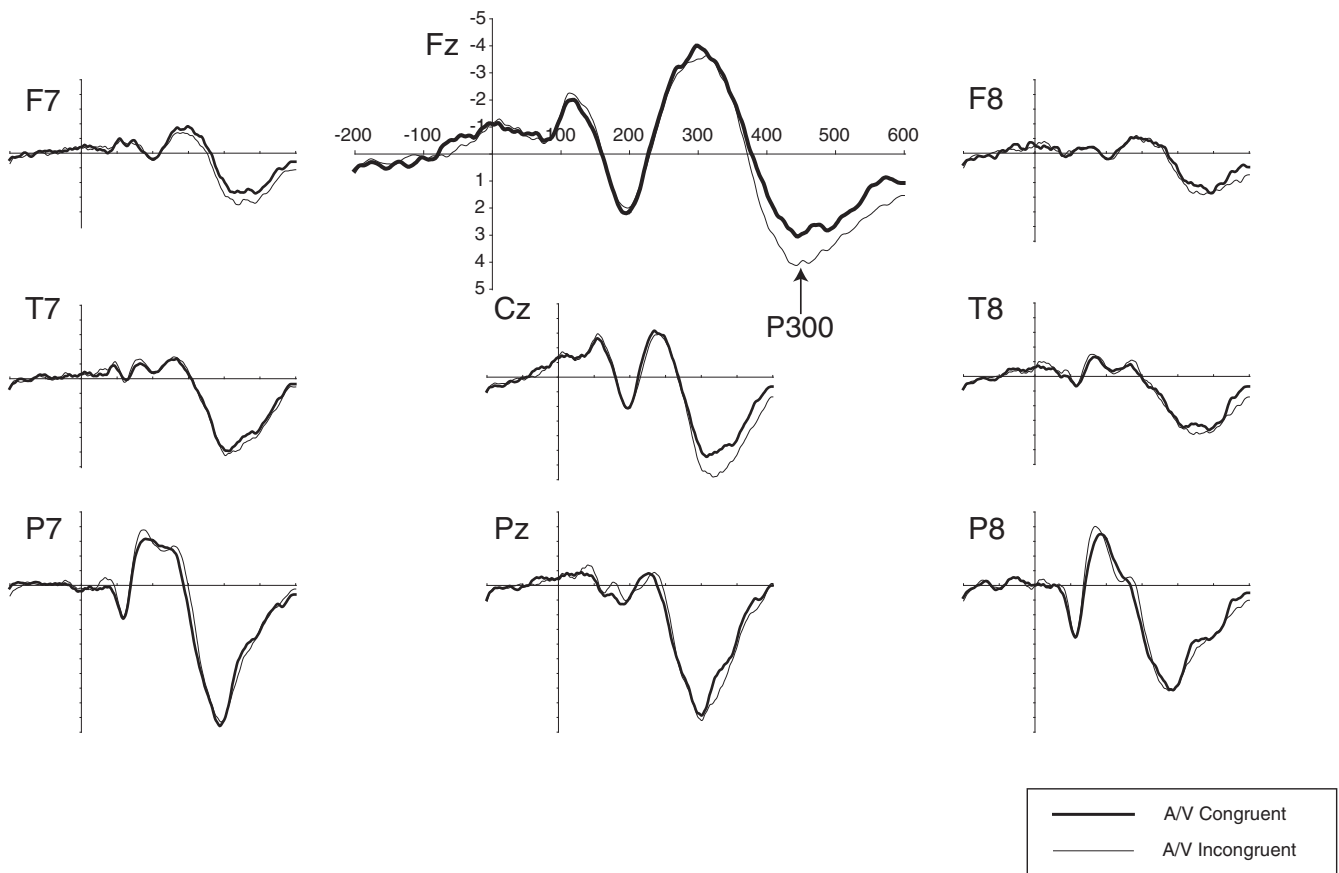

Fig. 3. MMN effect for congruent and incongruent trials. Subtraction waves (standards minus deviants) are presented separately for congruent and incongruent trials. Scalp map illustrates the electrode distribution of the contrast between congruent and incongruent MMNs.
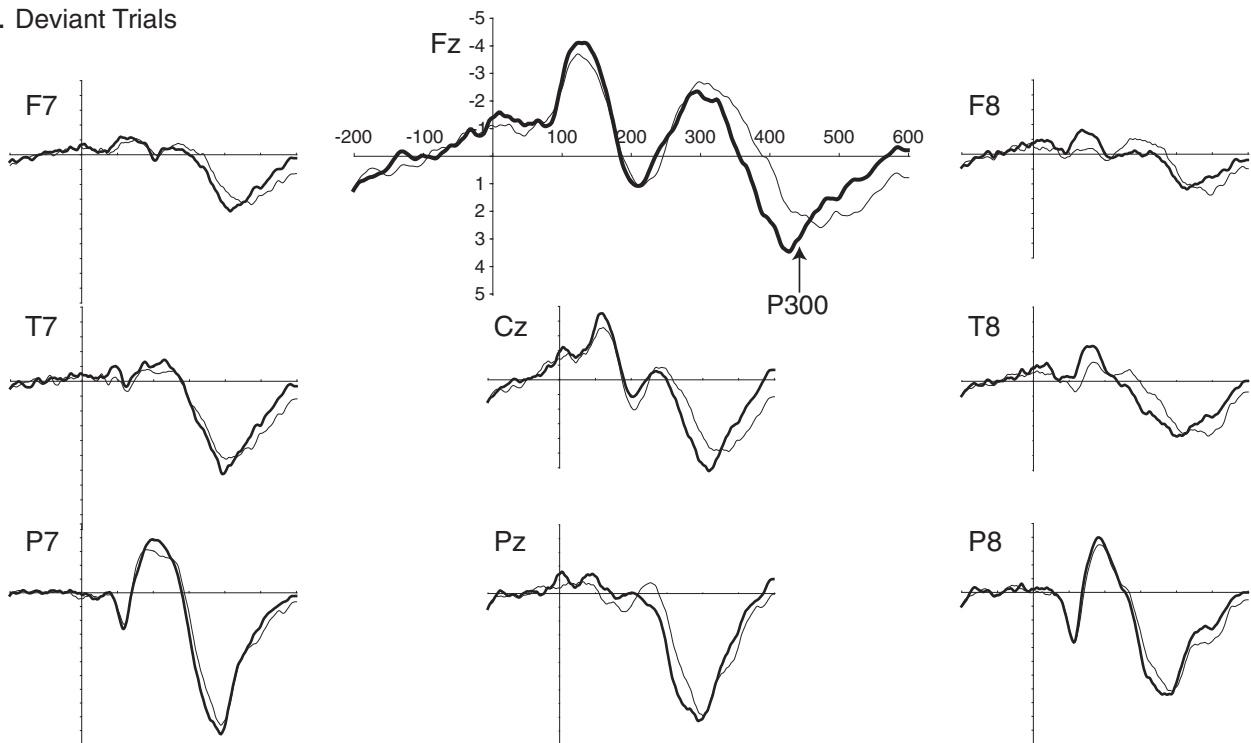
A. Standard Trials

B. Deviant Trials

Fig. 4. Effect of audiovisual (AV) congruency on P300s in (A) standard and (B) deviant trials.

This finding is somewhat consistent with that observed by Froyen et al. (2008), who found enhanced MMN when subjects viewed a repeated letter stimulus, compared to when this letter was presented asynchronously or not at all. That said, this earlier study did not directly compare the influence of congruent and incongruent letter stimuli. Rather, the letter stimulus remained the same

across all trials ('0'), and was congruent with the standard, such that when a deviant was played, it enhanced the auditory mismatch effect. The present study extends this finding in several key ways. First, letter stimuli changed on a trial-by-trial basis, which allowed us to present the two critical vowels 'E' and 'O' concurrent with either congruent or incongruent stimuli. Moreover, standard and deviant sounds were counterbalanced across runs within subjects, meaning that the /i/ and /o/ sounds were heard as both standards and deviants. These manipulations allowed us to determine that the observed modulation of the MMN is strictly dependent on the congruency of the letter with the auditory stimulus being played, rather than owing simply to the presence of a concurrent letter stimulus. Likewise, counterbalancing presentation of the key auditory and visual vowel stimuli meant that condition-wise differences cannot be due to vagaries in how the perceptual system processes the specific vowels or letters.

A recent study examined the effect of AV congruency during a visual analogue to the MMN paradigm (Froyen et al., 2010). Subjects viewed a stream of repeated letters that was interspersed with occasional oddball items. The authors found that playing a concurrent speech stimulus that was congruent with the repeated letter (standards) did not modulate the visual MMN observed for the deviant letter. They concluded that auditory speech inputs do not appear to influence early visual sensory processing. The present results are compatible with this interpretation, supporting the view that AV interactions for letters and speech sounds occur within early-going auditory sensory mechanisms. This is also compatible with tomographic studies showing specific AV interactions for letters and speech within posterior portions of the superior temporal lobe, but not within occipital or occipito-temporal brain regions (Herdman et al., 2006; Van Atteveldt et al., 2004).

In addition to addressing the timing of AV integration, examining the effects of congruent and incongruent presentations of visual information on the MMN response has implications for another conflict in the literature. Some studies suggest that the MMN response is contingent on a change in auditory stimuli and is unaffected by the demands of a simultaneous visual task (see Muller-Gass et al., 2006; Picton et al., 2000). On the other hand, more recent studies have suggested that MMN responses to speech can be modulated by concurrent visual articulatory stimuli (e.g., Saint-Amour et al., 2007; Sams et al., 1991). Such findings suggest that the MMN response to speech stimuli can be influenced by visual information such as articulatory gestures. However such findings leave open the question of whether the same is true in the case of letter–sound congruency. In the current study, congruent and incongruent visual information were presented with identical auditory stimuli in the congruent and incongruent oddball conditions. MMN responses are typically inter-preted as reflecting a comparison process between the deviant stimulus and an auditory short-term memory trace created by the repetition of the standard. The observed changes in the MMN response suggest that visual input interferes with this comparison process by modulating how the deviant is initially perceived, and as a consequence, interfering with the process of matching it to the repeated standard.

The results also appear to be compatible with a recent fMRI study in which Dutch-speaking subjects performed an auditory speech sound identification task in the context of task irrelevant visually presented congruent vs. incongruent letters (Blau et al., 2008). This study found that neural activation in superior temporal cortex was significantly increased in the context of congruent letter stimuli, compared to when incongruent letters were presented. These data again support the conclusion that early-going auditory speech perception mechanisms can be influenced in subtle ways by visual inputs.

The second set of analyses examined the influence of AV congruency on the P300 component, an index of visual attention that is known to be influenced by difficulty in response selection. As expected, congruency effects were observed in this component, for both standard and deviant auditory mismatch trials. This result suggests that visual attention to orthographic stimuli is influenced by concurrent presentation of a corresponding auditory stimulus. Interestingly, however, several prior studies have employed tasks in which individuals were asked to actively monitor both the auditory and visual streams (e.g., Herdman et al., 2006); as van Atteveldt et al. (2007) observed, actively monitoring for AV congruency can limit the observed neural effect compared to passive tasks. In the present study, this effect occurred under conditions where the auditory stream was not being actively monitored. Furthermore, the SOA of the auditory stimuli differed from that of the visual task, such that the two only coincided in a minority of trials (fewer than 15% of trials). Overall, the present study showed that visual letter detection performance is modulated by concurrent auditory information, even when individuals' attention is specifically directed away from the auditory stimuli.

Interestingly, the influence of congruency somewhat differed depending on the auditory mismatch manipulation. For standards, we found that incongruency led to increased P300 amplitudes; in contrast, for deviants incongruency yielded increased P300 latency effects. One interpretation is that increased amplitude of the P300 reflects the increased allocation of attentional resources toward a visual task (e.g. Wickens et al., 1977). Notably however, the actual visual stimuli were identical in both the congruent and incongruent conditions. As such, the differential effect appears to be due to the congruency of the auditory stimulus; in short, congruency appears to attenuate the cognitive processing required to categorize the visual letter stimulus as a vowel.

Also of interest is the observation that visual congruency had a different effect in the context of auditory deviants. Here we observed longer P300 latencies for incongruent stimuli. This effect is attributed not to the visual stimulus itself, since this was identical across the two trial types, but rather to the congruency of the visual letter and the unattended auditory vowel sound. Latency differences like these have previously been proposed to reflect delayed perceptual processing and categorization (Kutas et al., 1977).

There was an interesting relationship between the behavioral and ERP data. Subjects' accuracy was very high across all trial types, and was unaffected by either the congruency or mismatch manipulation. However, we did observe small RT differences, best characterized as faster responses for deviant congruent than deviant incongruent trials.[1] This pattern is quite consistent with the P300 latency data, which also showed longer latencies for the deviant incongruent vs. congruent stimuli. This observation is not altogether surprising, given prior studies showing shorter latency P300s on trials with shorter RTs (e.g., Holm et al., 2006). On the other hand, the MMN effects did not appear to be reflected in the RT data. That is, the strongest magnitude MMN effect was for standards vs. deviants, especially in congruent trials. The MMN effects were not closely paralleled by the behavioral data, in that we did not observe slower or less accurate responses on auditory mismatch trials. This seems to suggest that AV interaction effects on MMNs occurred at an early, pre-attentive stage of processing, and in a way that did not have cascading effects onto later stages of visual processing related to letter detection.

---

[1] Note that RTs for the standard incongruent trials were also significantly longer than for the standard congruent trials, however we did not observe significantly different P300 peak latencies in this case. The reason for this could be that the magnitude of the behavioral effect was much smaller for standard trials (6 ms, vs. 28 ms for the deviant trials). The P300 peak detection measure might not be sufficiently sensitive to detect a difference of this magnitude, due to the lower sample rate used to collect our EEG data, and the filtering and artifact rejection procedures used during preprocessing.

Overall, the discovery of a significant effect of congruency in both the MMN and P300 components supports the view that AV integration of letters and sounds occurs relatively early in the perceptual stream, and has a sustained effect during perceptual processing. This seems somewhat incompatible with an earlier view of AV integration that suggests simultaneously presented auditory and visual stimuli are processed independently in their respective sensory cortices and are integrated later in sensory processing (see Foxe and Schroeder, 2005 for review). For instance, Raij et al. (2000) found MEG differences on the order of 465 ± 50 ms post stimulus, which is consistent with the timing of the P300 modulations observed in this study, but much later than the MMN modulations we found. Our findings seem more in line with those of Herdman et al. (2006) for Japanese sounds and their corresponding Hiragana units, and with prior studies finding modulations in the MMN component when individuals view incongruent letter stimuli (Froyen et al., 2008) or oral-articulations (Saint-Amour et al., 2007).

In the present study, we chose to examine the integration of letters with their names, rather than letters and sounds. This differs somewhat from some earlier studies of letter–sound integration in languages with more consistent orthographies. In contrast, English letters rarely correspond to a unique sound; rather, an individual letter can map to several different phonemes depending on the context. So for instance the letter "A" produces the sound /e/ in GAVE and CAPE, but also /ae/ in HAVE and CAP, /a/ in CLAW, and so on. Indeed, English spelling is an outlier of sorts, marked by a high degree of irregularity compared to other alphabetic orthographies. Thus the letter–sound congruency effect we observed might be due to the match of the letter with the sound it maps to, or to the letter name itself. Thus, it remains an outstanding question whether it is possible to observe pure grapheme–phoneme congruency effects (e.g., A − /ae/) in English, similar to what has been previously found in highly transparent orthographies (Dutch: van Atteveldt et al., 2004; Japanese Hiragana: Herdman et al., 2006).

5. Conclusions

There is a growing body of literature aimed at understanding the role of AV integration on the perception of written and spoken language. This was addressed in the present study by examining how letter–sound congruency influences the MMN ERP component, which we argue indexes early-going and automatic auditory processing, vs. the P300, which is known to be influenced by cognitive factors related to attention, categorization and response selection. Results of the present study suggest that audiovisual integration of letters occurs relatively early in processing, in a period as early as 180 ms following the onset of the concurrently presented letter and sound, and that the influences of AV integration continue to affect later stages of processing.

Overall, the current study has implications for our understanding of adult perception of audiovisual letter stimuli, as well as contributing to our understanding of the nature of the MMN response. An understanding of the timing, or order, of processing of audiovisual stimuli in the brain provides valuable information as to the mechanisms by which skilled reading abilities are acquired and maintained in typically developing individuals. Moreover, the paradigm employed may prove to be useful in future studies examining perception and reading in younger populations, and in populations with reading difficulties. For example, if dyslexia stems from difficulties in integrating letters across the auditory and visual domains, it is important to know at what level auditory and visual letter information are integrated in normal readers in order to infer the level at which processing may have been disrupted in individuals with dyslexia. For instance, there is evidence that AV congruency effects on MMN amplitude increase over the course of normal reading development (Froyen et al., 2009). Recent evidence also suggests that individuals with dyslexia show a persistent pattern of insensitivity to AV congruency, which might relate to their underlying difficulties with pairing written and oral language (Blau et al., 2009, 2010).

References

Blau, V., van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2008. Task-irrelevant visual letters interact with the processing of speech sounds in heteromodal and unimodal cortex. The European Journal of Neuroscience 28, 500–509.

Blau, V., van Atteveldt, N., Ekkebus, M., Goebel, R., Blomert, L., 2009. Reduced neural integration of letters and speech sounds links phonological and reading deficits in adult dyslexia. Current Biology 19, 503–508.

Blau, V., Reithler, J., van Atteveldt, N., Seitz, J., Gerretsen, P., Goebels, R., Blomert, L., 2010. Deviant processing of letters and speech sounds as proximate cause of reading failure: a functional magnetic resonance imaging study of dyslexic children. Brain 133, 868–879.

Foxe, J.J., Schroeder, C.E., 2005. The case for feedforward multisensory convergence during early cortical processing. NeuroReport 16, 419–423.

Froyen, D.J.W., van Atteveldt, N., Bonte, M.L., Blomert, L., 2008. Cross-modal enhancement indicates early automatic integration of letters and speech-sounds. Neuroscience Letters 430, 23–28.

Froyen, D.J.W., Bonte, M.L., van Atteveldt, N., Blomert, L., 2009. The long road to automation: neurocognitive development of letter-speech sound processing. Journal of Cognitive Neuroscience 21 (3), 567–580.

Froyen, D., van Atteveldt, N., Blomert, L., 2010. Exploring the role of low level visual processing in letter-speech sound integration: a visual MMN study. Frontiers in Integrative Neuroscience 13, 1–9.

Herdman, A.T., Fujioka, T., Chau, W., Ross, B., Pantev, C., Picton, T.W., 2006. Cortical oscillations related to processing congruent and incongruent grapheme–phoneme pairs. Neuroscience Letters 399, 61–66.

Holm, A., Ranta-aho, P.O., Sallinen, M., Karjalainen, P.A., Müller, K., 2006. Relationship of P300 single-trial responses with reaction time and preceding stimulus sequence. International Journal of Psychophysiology 61, 244–252.

Kutas, M., McCarthy, G., Donchin, E., 1977. Augmenting mental chronometry: the P300 as a measure of stimulus evaluation time. Science 197 (4305), 792–795.

Laurienti, P.J., Wallace, M.T., Maldjian, J.A., Susi, C.M., Stein, B.E., Burdette, J.H., 2003. Cross-modal sensory processing in the anterior cingulate and medial prefrontal cortices. Human Brain Mapping 19, 213–223.

Magnan, A., Ecalle, J., Veuillet, E., Collet, L., 2004. The effects of an audio–visual training program in dyslexic children. Dyslexia 10, 131–140.

McGurk, H., MacDonald, J., 1976. Hearing lips and seeing voices. Nature 264, 746–748.

Molholm, S., Ritter, W., Murray, M.M., Javitt, D.C., Schroeder, C.E., Foxe, J.J., 2002. Multisensory auditory–visual interactions during early sensory processing in humans: a high-density electrical mapping study. Cognitive Brain Research 14, 115–128.

Muller-Gass, A., Stelmack, R.M., Campbell, K.B., 2006. The effect of visual task difficulty and attentional direction on the detection of acoustic change as indexed by the mismatch negativity. Brain Research 1078, 112–130.

Näätänen, R., 2000. Mismatch negativity (MMN): perspectives for application. International Journal of Psychophysiology 3, 3–10.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clinical Neurophysiology 118, 2544–2590.

Picton, T.W., Alain, C., Otten, L., Ritter, W., Achim, A., 2000. Mismatch negativity: different water in the same river. Audiology & Neuro-Otology 5, 111–139.

Polich, J., 2007. Updating p300: an integrative theory of P3a and P3b. Clinical Neurophysiology 118 (10), 2128–2148.

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

Saint-Amour, D., De Sanctis, P., Molholm, S., Ritter, W., Foxe, J.J., 2007. Seeing voices: high-density electrical mapping and source-analysis of the multisensory mismatch negativity evoked during the McGurk illusion. Neuropsychologia 45, 587–597.

Sams, M., Aulanko, R., Hamalainen, M., Hari, R., Lounasmaa, O.V., Lu, S.T., et al., 1991. Seeing speech: visual information from lip movements modifies activities in the human auditory cortex. Neuroscience Letters 127, 141–145.

Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. The Journal of the Acoustical Society of America 26, 212–215.

Teder-Sälejärvi, W.A., McDonald, J.J., Di Russo, F., Hillyard, S.A., 2002. An analysis of audio–visual crossmodal integration by means of event-related potential (ERP) recordings. Cognitive Brain Research 14 (1), 106–114.

van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2004. Integration of letters and speech sounds in the human brain. Neuron 43, 271–282.

Van Atteveldt, N., Formisano, E., Goebel, R., Blomert, L., 2007. Top-down task effects overrule automatic multisensory responses to letter-sound pairs in auditory association cortex. Neuroimage 36, 1345–1360.

Wickens, C.D., Israel, J., Donchin, E., 1977. The event related control potential as an index of task workload in adaptive man-machine systems. Proceedings of the 21st Annual Meeting of the Human Factors Society, USA.

Widmann, A., Kujala, T., Tervaniemi, M., Kujala, A., Schröger, E., 2004. From symbols to sounds: visual symbolic information activates sound representations. Psychophysiology 41, 709–715.