ELSEVIER

Cognitive Brain Research 14 (2002) 106–114

COGNITIVE BRAIN RESEARCH

www.elsevier.com/locate/bres

Research report

An analysis of audio-visual crossmodal integration by means of event-related potential (ERP) recordings

W.A. Teder-Sälejärvi[a,*], J.J. McDonald[b], F. Di Russo[a], S.A. Hillyard[a]

[a]*Department of Neurosciences 0608, School of Medicine, University of California–San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0608, USA*
[b]*Simon-Fraser University, Vancouver, Canada*

Accepted 20 July 2001

Abstract

Crossmodal integration was studied in humans by presenting random sequences of auditory (brief noise bursts), visual (flashes), and audiovisual (simultaneous noise bursts and flashes) stimuli from a central location at irregular intervals between 600 and 800 ms. The subjects' task was to press a button to infrequent and unpredictable ($P=0.15$) target stimuli that could be either a more intense noise burst, a brighter flash, or a combination of the two. In accordance with previous studies, behavioral data showed that bimodal target stimuli were responded to much faster and were identified more accurately than the unimodal target stimuli. The neural basis of this crossmodal interaction was investigated by subtracting the ERPs to the auditory (A) and the visual (V) stimuli alone from the ERP to the combined audiovisual (AV) stimuli (i.e. interaction=AV−(A+V)). Using this approach, we replicated previous reports of both early (at around 40 ms) and late (after 100 ms) ERP interaction effects. However, it appears that the very early interaction effects can be largely accounted for by an anticipatory ERP that precedes both the unimodal and bimodal stimuli. In calculating the ERP interaction this slow shift is subtracted twice, resulting in an apparent shift of the opposite polarity that may be confounded with actual crossmodal interactions.

Theme: Neural basis of behavior

Topic: Cognition

Keywords: Auditory attention; Visual attention; Crossmodal attention; Multimodal attention

1. Introduction

The ability to integrate stimuli in different modalities to form unified percepts is a fundamental component of sensory guided behavior and cognition. Psychophysical studies have shown that behavioral responses to multimodal inputs presented in close spatial and temporal proximity are typically faster and more accurate than those made to the unimodal stimuli alone [19,15]. The neuronal bases for this cross-modal enhancement of perception and behavior is being actively investigated in both animals and humans (reviewed in Refs. [17,3]), but our understanding of the underlying mechanisms is still at a preliminary level. A common paradigm for demonstrating cross-modal facilitation has been to compare neural responses to each of two unimodal stimuli presented alone with the response to the bimodal combination of the two [17]. If the bimodal response is greater than the sum of the individual unimodal responses, it may be concluded that some form of facilitatory intermodal integration is taking place. Using this type of design, evidence for cross-modal enhancement has been demonstrated in neurophysiological recordings from multisensory neurons in a variety of brain structures including the superior colliculus, posterior parietal cortex, superior temporal cortex, and neurostriatum [5] (reviewed in Ref. [17]).

In humans, cross-modal enhancement of neural activity has been demonstrated using both electrophysiological recordings of event-related brain potentials [6,5] and neuroimaging [3] measures. One of the advantages of ERP recordings is that the precise time course of cross-modal integration can be studied in different brain areas. How-

*Corresponding author. Tel.: +1-858-534-1389; fax: +1-858-534-5562.

E-mail address: wat@sdepl.ucsd.edu (W.A. Teder-Sälejärvi).

0926-6410/02/$ – see front matter
PII: S0926-6410(02)00065-4

ever, a potential problem may arise when subjects are required to attend to the stimuli in the bimodal–unimodal comparison paradigm. For example, in a study of auditory (A)/visual (V) integration, evidence for cross-modal interaction would be obtained when the amplitude of the bimodal (AV) ERP differs from the sum of the individual unimodal ERPs (A+V). Such an interaction would be evident as voltage deflections in the AV−(A+V) difference ERP. The problem comes in if anticipatory slow potentials are elicited in association with each of the three types of stimulus events (A, V, and AV). Such potentials would be added only once but subtracted twice in the AV−(A+V) difference wave, creating a deflection that might be mistaken for a true auditory–visual interaction. For example, an anticipatory ramp-like negativity (such as the CNV [18]) that arose before each stimulus and continued for a time after stimulus onset might appear as a ramp-like positivity in the AV−(A+V) difference wave, due to its having been subtracted twice. Depending on how the difference wave was baselined, such an anticipatory slow wave could appear as a significant voltage deflection at a very short latency after stimulus onset. The principal aim of the present study was to separate the possible influences of anticipatory slow waves from true cross-modal interactions in processing as reflected in the AV−(A+V) difference wave. This separation was achieved by showing, first, that slow wave deflections in the difference wave became significant at shorter latencies when earlier baseline periods were chosen for the measurements and, second, that dipole modeling can distinguish the sources of these early slow-wave effects from later, genuine cross-modal interactions. In addition, phasic cross-modal interactions were separated from anticipatory slow potentials by means of high-pass filtering of the AV−(A+V) difference waves. A secondary aim was to estimate the locations of the brain areas in which cross-modal interactions take place using dipole modeling.

2. Methods

2.1. Participants

Fifteen healthy adults (eight female; ages 19–29 years, mean age 21.1 years) participated in this study after giving written informed consent. Each participant had normal or corrected-to-normal vision and was tested in the laboratory to confirm normal hearing.

2.2. Stimuli and apparatus

The experiment was conducted in a sound-attenuated chamber with a background sound level of 32 dB (A) and a background luminance of 2 cd/m^2. Participants sat in the chamber and faced a loudspeaker with a red light-emitting diode (LED) mounted in its cone. Participants were instructed to focus their eyes on the central loudspeaker throughout each experimental run and to press a button to infrequent target stimuli in either the auditory or visual modality.

Randomized stimulus sequences were presented consisting of equiprobable auditory stimuli (500–15 000 Hz 'pink' noise bursts, at 76 dB SPL), visual stimuli (LED flashes, 75 cd/m^2), and audiovisual stimuli (a simultaneous combination of noise burst and LED flash) at irregular intervals of 600–800 ms. All stimuli were of 33 ms duration, and infrequent target stimuli (P=0.15) presented in both modalities had a higher intensity than the frequent standard stimuli. The auditory target stimuli were noise bursts of slightly increased loudness (by 3–9 dB), the visual targets were flashes with slightly increased brightness. The audiovisual target stimuli were a combination of the two targets. In an initial practice run, the task difficulty (i.e. target discriminability) was individually adjusted to about 75% correct responses for both auditory and visual target stimuli.

Each volunteer participated in one 1.5-h session that consisted of an initial practice run and 16 experimental runs. Between runs, participants rested for 1–5 min. Each run consisted of 300 stimuli. Thus, each subject received a total of 4800 stimuli. All procedures were approved by the UCSD human subjects review committee.

2.3. EEG recording and data analysis

Electroencephalographic signals were recorded from 64 tin electrodes, including 56 sites from the 10-10 system (FPz, FP1, FP2, Fz, F1, F2, F3, F4, F5, F6, F7, F8, FCz, FC1, FC2, FC3, FC4, FC5, FC6, Cz, C1, C2, C3, C4, C5, C6, T7, T8, CPz, CP1, CP2, CP3, CP4, CP5, CP6, Pz, P1, P2, P3, P4, P5, P6, P7, P8, P9, P10, POz, PO3, PO4, PO7, PO8, Oz, O1, O2, Iz, and A1) [1], and four additional electrodes located inferior to the occipital row of electrodes. Horizontal electrooculographic (EOG) signals were recorded bipolarly using electrodes at the left and right external canthi, and vertical EOG signals were recorded from an electrode below the left eye. All scalp electrodes, as well as the electrode below the left eye, were referenced to an electrode on the right mastoid (A2). Electrode impedances were kept below 5 kΩ.

All signals were amplified with a gain of 20 000 and a bandpass of 0.1–100 Hz (−12 dB/octave; 3 dB attenuation), digitized at a rate of 250 Hz, and stored on disk for off-line averaging. Automated artifact rejection was performed prior to averaging to discard trials during which an eye movement, blink, or amplifier blocking occurred. Signals from the remaining trials were averaged in 3000-ms epochs that started 1000 ms before each stimulus. The averages were digitally low-pass filtered with a Gaussian finite impulse function (3 dB attenuation at 46 Hz) to remove high-frequency noise produced by muscle movements and external electrical sources. The filtered averages

were digitally re-referenced to the average of the mastoids. Cross-modal interaction was investigated by subtracting the ERPs to the auditory (A) and the visual (V) stimuli alone from the ERP to the combined audiovisual (AV) stimuli (i.e. interaction=AV−(A+V). Only ERPs to the standard (non-target) stimuli were included in this analysis.

In a separate analysis, the averaged waveforms were digitally high-pass filtered with a Gaussian finite impulse function (3 dB attenuation at 2 Hz) in order to minimize the contribution of pre- and post-stimulus slow potentials. By using a peak-detection algorithm, several positive and negative peaks were identified in the AV−(A+V) difference waveform between 80 and about 340 ms. For statistical testing two-tailed *t*-tests were carried out comparing mean amplitudes within specified time windows that included the peaks against the −100 to 0 ms pre-stimulus baseline. Amplitudes were measured at either frontal (F1, Fz, F2), central (C1, Cz, C2), or parietal (P1, Pz, P2) sites. These sites coincided with the regions of maximum amplitude of the various peaks. The following time-windows were tested: 80–88, 98–112, 128–136, 160–172, 242–266, and 318–334 ms.

2.4. Source localization

Estimation of the dipolar sources of ERP components was carried out using Brain Electrical Source Analysis (BESA version 2.1). The BESA algorithm estimates the location and the orientation of multiple equivalent dipolar sources by calculating the scalp distribution that would be obtained for a given dipole model (forward solution) and comparing it to the original ERP distribution. Interactive changes in the location and in the orientation in the dipole sources lead to minimization of the residual variance (RV) between the model and the observed spatio-temporal ERP distribution [16]. The energy criterion of the BESA was set at 15% to reduce the interaction among dipoles, and the separation criterion was set at 10% to optimize the separation of the source waveforms that differ over time. In these calculations, BESA assumed an idealized three-shell spherical head model with a radius of 85 mm and scalp and skull thicknesses of 6 and 7 mm, respectively. Dipoles pairs were fit sequentially over specific latency ranges (given below) to correspond with the distinctive components in the waveform.

The approach for spatio-temporal dipole modeling was to fit symmetrical pairs of dipoles (constrained to be mirror-symmetrical in location and in orientation) over restricted time intervals corresponding to the components of interest in the multi-modal interaction (AV−(A+V)) difference-wave. Dipole pairs were chosen because the scalp voltage and current density distributions of the interaction components were essentially bilaterally symmetrical, indicative of mirror-image sources in the two hemispheres [16]. To account for the slow deflection having a very early onset in the unfiltered waveforms, a pair of dipoles was fit over the time interval of 40–100 ms (dipoles ≠1–2). A second pair of dipoles was then added to the model and fit over the time interval of 130–170 ms to account for the first major phasic deflection in the interaction waveform in this latency range. Finally, a third pair was fit over the interval 170–230 ms to account for the second major phasic deflection. The dipole models obtained were little affected by using different symmetry constraints or different starting locations in the dipole fitting.

Dipole modeling was also carried out on the high-pass filtered AV−(A+V) waveforms, taking the model that was fit to the unfiltered waveforms as a starting point. The dipole pairs were then fit sequentially in both location and orientation to the major deflections in the filtered difference wave in the following order: dipoles ≠3–4 (130–170 ms), dipoles ≠4–5 (170–230 ms), dipoles ≠1–2 (280–340 ms).

3. Results

3.1. Behavioral data

3.1.1. Accuracy

Overall, only about 60% of the auditory targets were correctly responded to during the experimental runs, a significant difference compared to about 85% correctly identified visual targets ($t[14]=6.17$, $P<0.0001$). The bimodal targets were correctly identified at a significantly higher rate (92%) than either the auditory targets ($t[14]=6.25$, $P<0.0001$), or the visual targets alone ($t[14]=3.48$, $P<0.004$). False alarm rates were very low, 2.52% for auditory, 2.89% for visual, and 6.56% for audiovisual stimuli.[1]

3.1.2. Reaction time (RT)

The RTs for visual targets (mean=464 ms, S.E.M. 8.5 ms) were shorter than the RTs to auditory targets (mean=495 ms, S.E.M. 8.5 ms). This difference was statistically significant ($t[14]=4.04$, $P<0.002$). However, RTs to the bimodal targets (mean=401 ms, S.E.M. 12.1 ms) were markedly facilitated when compared to either auditory ($t[14]=15.94$, $P<0.0001$) or visual ($t[14]=8.52$, $P<0.0001$) targets.

3.2. Electrophysiological data

Fig. 1 shows the ERPs elicited by the unimodal and bimodal stimuli. The waveforms showed a typical morphology that included auditory and visual P1, N1, and P2

[1]Although target discriminability was initially equated for the auditory and visual stimuli in practice runs, it turned out that performance improved substantially for the visual targets during the experimental runs relative to the auditory targets. It is not clear why that occurred.

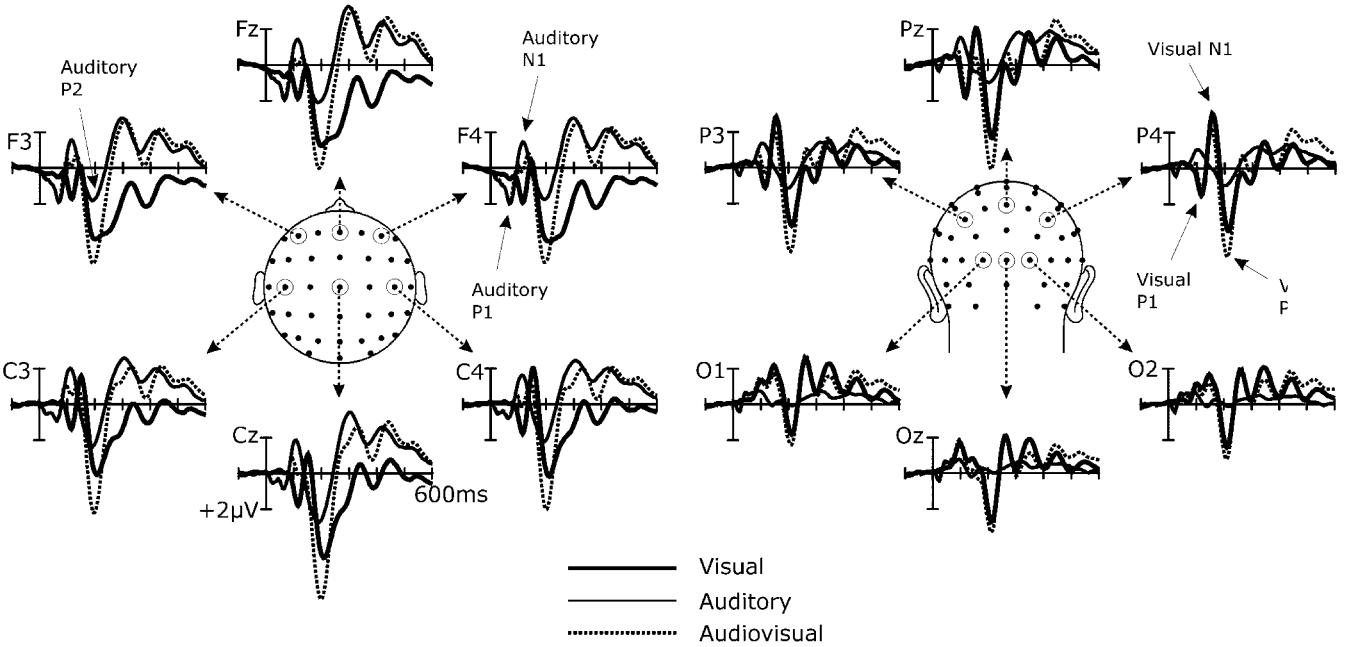

Fig. 1. Grand-average ERPs (n=15) elicited by unimodal auditory (A) and visual (V) stimuli and by bimodal (audiovisual, AV) stimuli at a subset of anterior (left) and posterior (right) electrodes.

components. However, it can be seen at frontal sites (left) that the stimulus-triggered ERPs were preceded by a slow ramping positivity that apparently began before stimulus presentation, which was paralleled by a slow negativity over parietal and occipital sites (right).

The thick traces in Fig. 2 show the cross-modal interaction waveform (AV−(A+V)) obtained by subtracting the ERPs to the auditory (A) and visual (V) unimodal stimuli from the ERP to the bimodal audiovisual (AV) stimuli. It can be seen that the bimodal response is not simply the linear sum of separately recorded unimodal activity. Over anterior sites (left) the interaction waveform began with a slow-ramping negativity followed by a positive–negative–positive complex of peaks during 100–160 ms and then a prominent negative deflection at about 250 ms. Over posterior sites (right), the interaction waveform included a

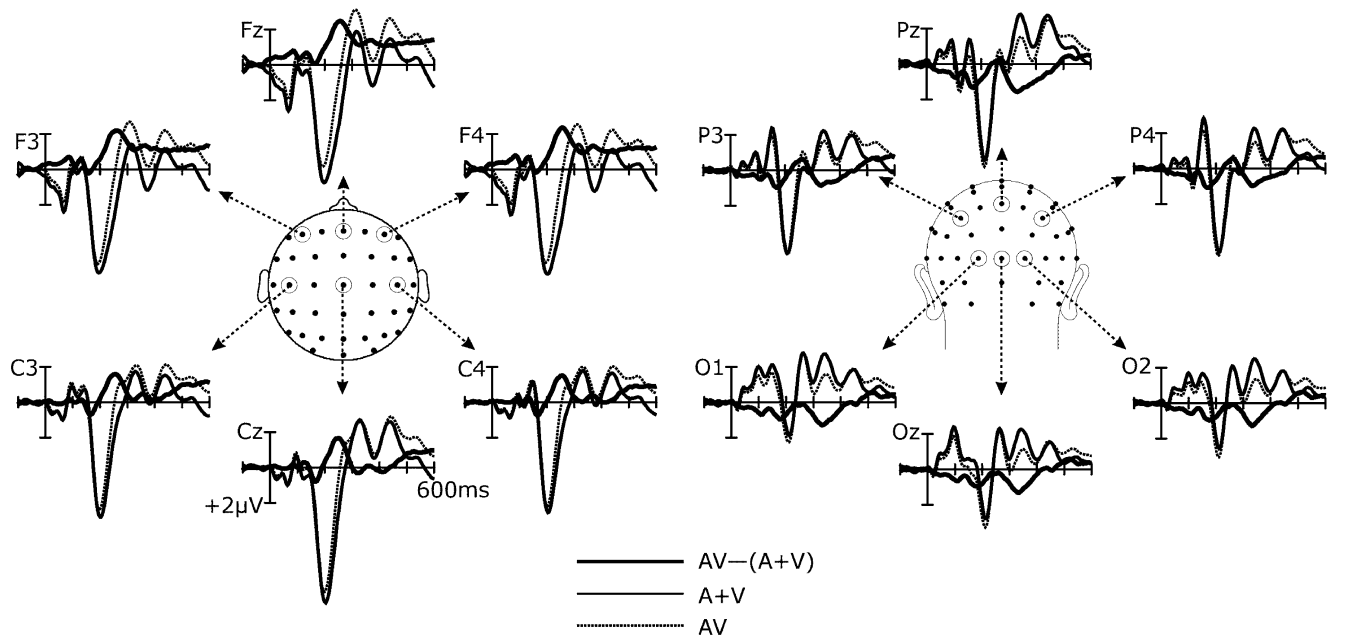

Fig. 2. Superimposition of grand-average ERPs (n=15) to bimodal (AV) stimuli, the algebraic sum of ERPs to unimodal auditory (A) and visual (V) stimuli, and the difference between the two waveforms (i.e. AV−(A+V)) at a subset of anterior (left) and posterior (right) electrodes.

slow positive shift that began at or before stimulus onset as well as some distinct peaks. The first peak was a small negativity peaking between 130 and 140 ms, followed by a positive peak at around 170 ms. The most prominent peak in the posterior interaction waveform was a relatively broad negativity peaking at around 250–270 ms.

To determine the latency at which the early slow deflections in the interaction waveform became significant, the mean amplitude over successive 10-ms intervals was tested with respect to a baseline voltage over the 100-ms interval prior to stimulus onset [6]. Fig. 3 shows that the slow deflection begins to have a well-defined topography, negative frontally and positive occipito-parietally at around 10–20 ms after stimulus onset. At frontal sites (mean of F1, Fz, and F2) the negative deviation from baseline first became significant at 20–30 ms after stimulus onset ($t[14]=2.14$, $P<0.05$), and at posterior sites (mean of P1, Pz, and P2) the positive deviation first became significant at 30–40 ms ($t[14]=3.12$, $P<0.01$). These deviations became increasingly larger and more significant over the first 100 ms of the post-stimulus period. However, when the amplitudes of these slow deflections were tested with respect to an earlier reference baseline of −100 to −50 ms, they became significant at even earlier time intervals. With this earlier baseline, the slow deviation over frontal sites first became significant at 10–20 ms ($t[14]=2.30$, $P<0.04$) and over posterior sites at 20–30 ms ($t[14]=2.43$, $P<0.03$). This increase in amplitude of the very early deviations with respect to the earlier baselining period would be expected if those deviations were consequences of a sloping pre-stimulus baseline.

The general similarity of the scalp topography of the cross-modal interaction wave (Fig. 3) between the very early phase (10–60 ms) and the late phase (320–360 ms) suggests that the several distinct peaks mentioned above were superimposed on a singular slow potential shift that persisted throughout the interval 0–360 ms post-stimulus. As described below, this suggestion was confirmed by analysis of high-pass filtered waveforms and by dipole modeling of the slow wave and superimposed peaks.

3.2.1. High-pass filtered waveforms

Fig. 4 shows the high-pass filtered difference waves (AV−(A+V)), in which the anticipatory slow potential shifts were greatly attenuated. Statistical tests of the phasic deflections in the filtered difference waves revealed that the earliest significant deviation from the −100 ms to zero baseline occurred in the 98–112-ms range, with increased

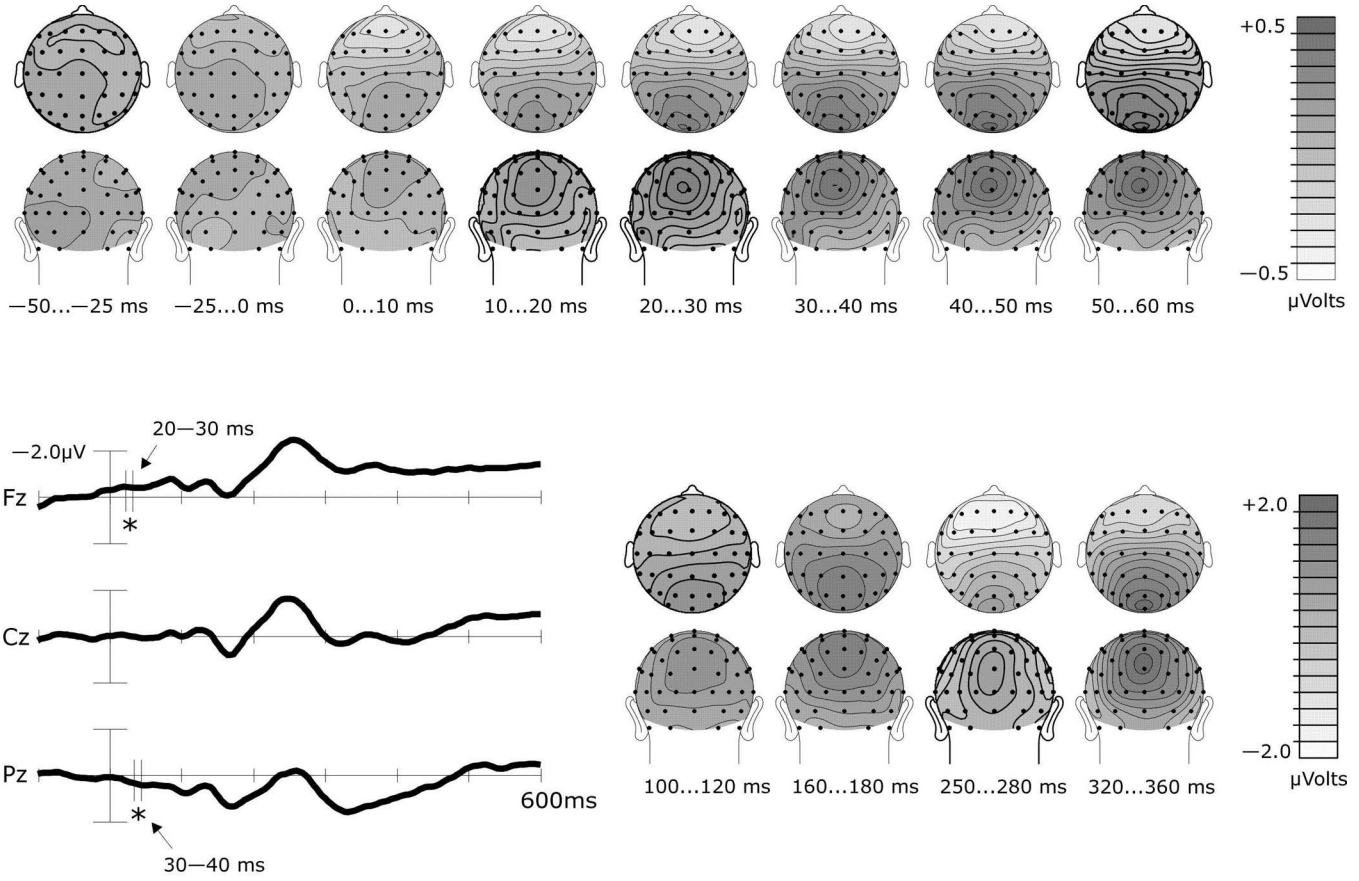

Fig. 3. Spline-interpolated isovoltage maps of the grand-average AV−(A+V) difference waveform at successive time intervals. The voltage measurements were obtained with respect to a −100–0 ms pre-stimulus baseline. Difference waveforms are shown for sites Fz, Cz, and Pz. The difference waveforms became significantly different from baseline at frontal sites during the time interval of 20–30 ms and at posterior sites during 30–40 ms.

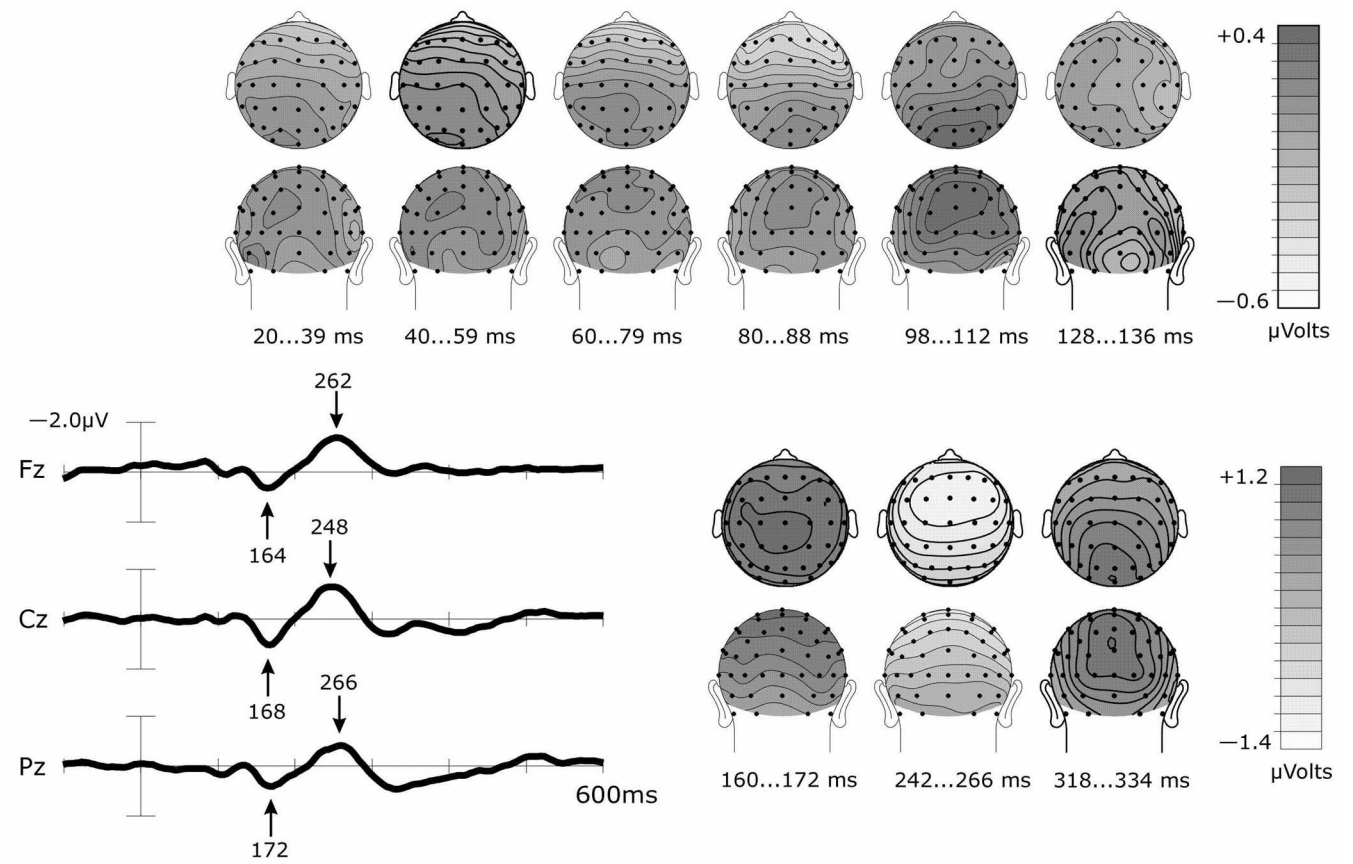

Fig. 4. Same as Fig. 3 for high-pass filtered waveforms. Note that overlapping slow potentials are greatly attenuated so that more veridical scalp distributions of the phasic deflections at 160–170 and 230–260 ms can be seen.

positivity at parietal (mean of P1, Pz, and P2) sites ($t[14]=2.20$, $P<0.05$). Considering that the scalp distribution of this positive deviation (Fig. 4) closely resembled that of the slow-wave deflection shown in Fig. 3, it seems likely that it represents a residual slow potential that was attenuated but not eliminated by the high-pass filtering. The first major deflection in the filtered difference waveform was a positivity peaking at 164–172 ms with an amplitude maximum that was highly significant ($t[14]=4.66$, $P<0.001$) at central (mean of C1, Cz, and C2) sites in the interval 160–172 ms. A subsequent negative deflection peaking at 250–260 ms with a frontal amplitude maximum was also highly significant ($t[14]=6.73$, $P<0.0001$) at frontal (mean of F1, Fz, and F2) sites in the interval 242–266 ms. A late slow positivity in the 300–400 ms range was also significant over posterior sites ($t[14]=4.98$, $P<0.001$ at parietal (mean of P1, Pz, and P2) sites in the interval 318–334 ms).

3.2.2. Dipole modeling

Inverse dipole modeling of the A−(A+V) difference wave was carried out on the grand average waveforms (both before and after high-pass filtering) using the BESA algorithm (Fig. 5). For the unfiltered difference waves the initial slow wave distribution was fit over 40–100 ms with a horizontally oriented dipole pair (≠1–2) located in the medial centro-parietal region. The source waveforms for this dipole pair suggest that slow activity arising from this same source persisted until at least 400 ms. The activity ranging from 130 to 170 ms was fit with a pair of vertically oriented dipoles (≠3–4) that were situated in ventral occipital cortex. The activity ranging from 170 to 230 ms was fit with a pair of dipoles (≠5–6) that were situated in the anterior temporal peri-sylvian cortex. This multi-dipole model accounted for more than 97% (2.8% residual variance) of the variance in scalp voltage topography over the time range 40–383 ms.

Fig. 5 also depicts the modified dipole model that was fit to the high-pass filtered difference waves (see Section 2). The dipole locations and source waveforms were highly similar to those of the unfiltered model and accounted for a similar amount of variance (96.8%) over the time range of the phasic deflections. The only appreciable difference between the models was the absence of the large slow potentials in the source waveforms of dipoles ≠1–2 in the model fit to the filtered waveforms.

The time periods over which the different sources contributed to the unfiltered difference wave distributions is depicted in terms of source maps in Fig. 6. At 60 ms post-stimulus only the centro-parietal dipole pair (≠1–2)

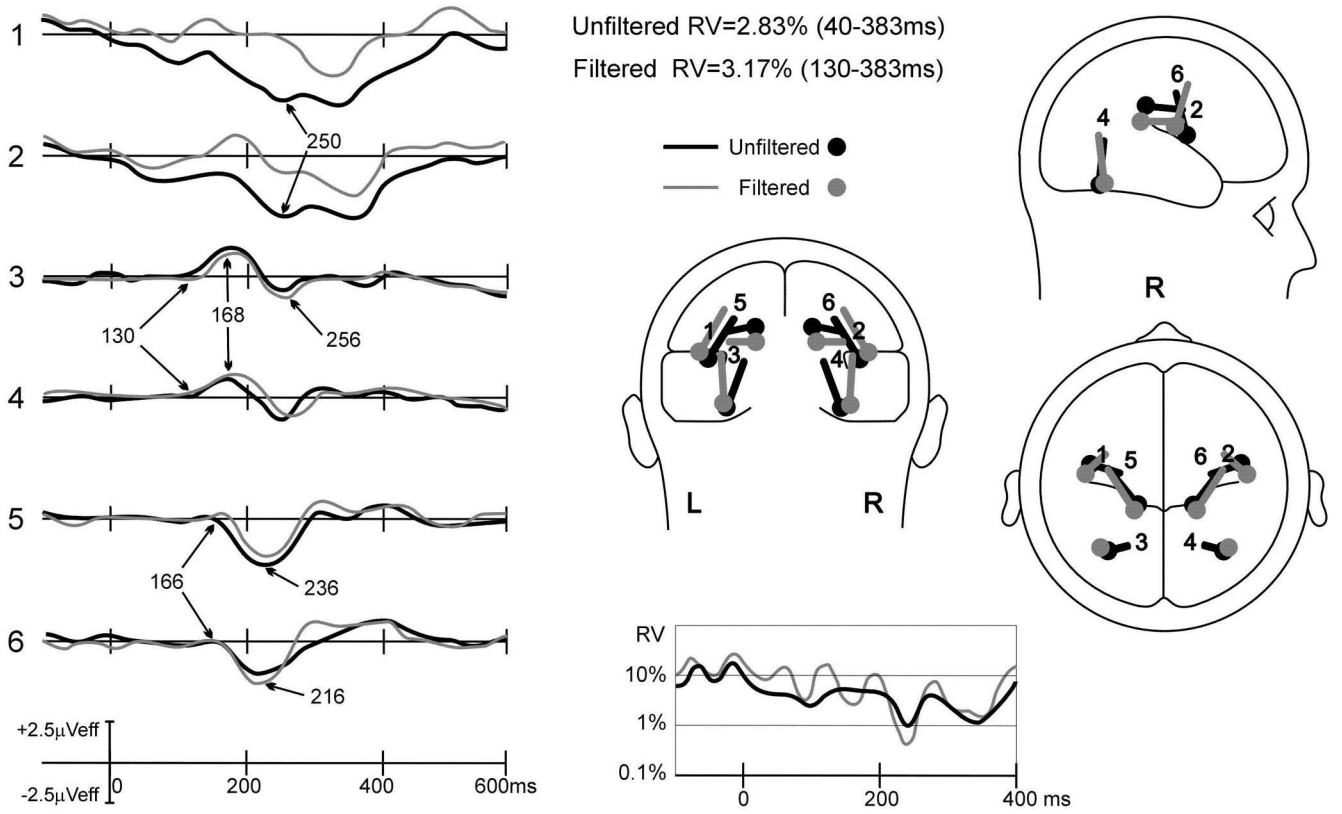

Fig. 5. BESA dipole model fit to grand-average AV−(A+V) difference waveforms for both filtered (gray) and unfiltered (black) data. Waveforms at the left show the time course of source activity for each of the modeled dipoles. Polarity convention for source waveforms: positivity at pointed end of dipole upwards.

was active, but this source continued to contribute activity to the model at each of the successive time points shown (160, 230, and 340 ms). In contrast, the ventral occipital dipole pair (≠3–4) was maximally active at around 160 ms, while the temporal peri-sylvian dipole contributed activity at around 230 ms.

4. Discussion

The present results provide evidence that slow potentials beginning prior to stimulus onset can contribute to the bimodal minus unimodal difference waveform (AV−(A+V)) that is often taken as an index of cross-modal interactions in neural processing. In this study, the slow anticipatory potentials were found to be positive over the anterior scalp and negative over the posterior scalp and probably belong to the CNV family of ERPs that precede perceptual decisions and discriminative responses [8]. Since these anticipatory ERPs were identical before each stimulus in the randomized sequences used here, these slow waves were added once but subtracted twice in the AV−(A+V) difference wave. The resulting difference wave deflections were evident as a slow, ramp-like negativity anteriorly and positivity posteriorly, both of which appeared to begin before stimulus onset. When these slow waves were baselined over the 100 ms interval before stimulus onset, the deviations from baseline became significantly different from zero at 20–30 ms post-stimulus anteriorly and 30–40 ms posteriorly. However, when the baseline was moved earlier to 100–50 ms prior to stimulus onset, these deflections became significant at 10–20 and 20–30 ms, respectively.

The finding that the deviations from baseline became significant at earlier post-stimulus intervals for earlier baseline periods strongly supports the conclusion that these effects were due to slow potentials that began before stimulus onset and extended into the post-stimulus period rather than to true cross-modal interactions. That is, an amplitude measurement taken at a particular time point along a sloping waveform will be larger when the baseline for measurement is moved to an earlier interval along that waveform. Moreover, it is difficult to understand how auditory–visual interactions in neural processing could occur in the brain prior to 40 ms, since the earliest evoked activity in primary visual cortex to moderate intensity stimuli has been reported to occur in the 45–55 ms range [9,4,10]. Accordingly, it would seem prudent to consider the possibility that very early deflections found in bimodal minus unimodal difference waves in previous studies [6] might be attributable, at least in part, to anticipatory slow potentials of the type observed here. It should be noted that

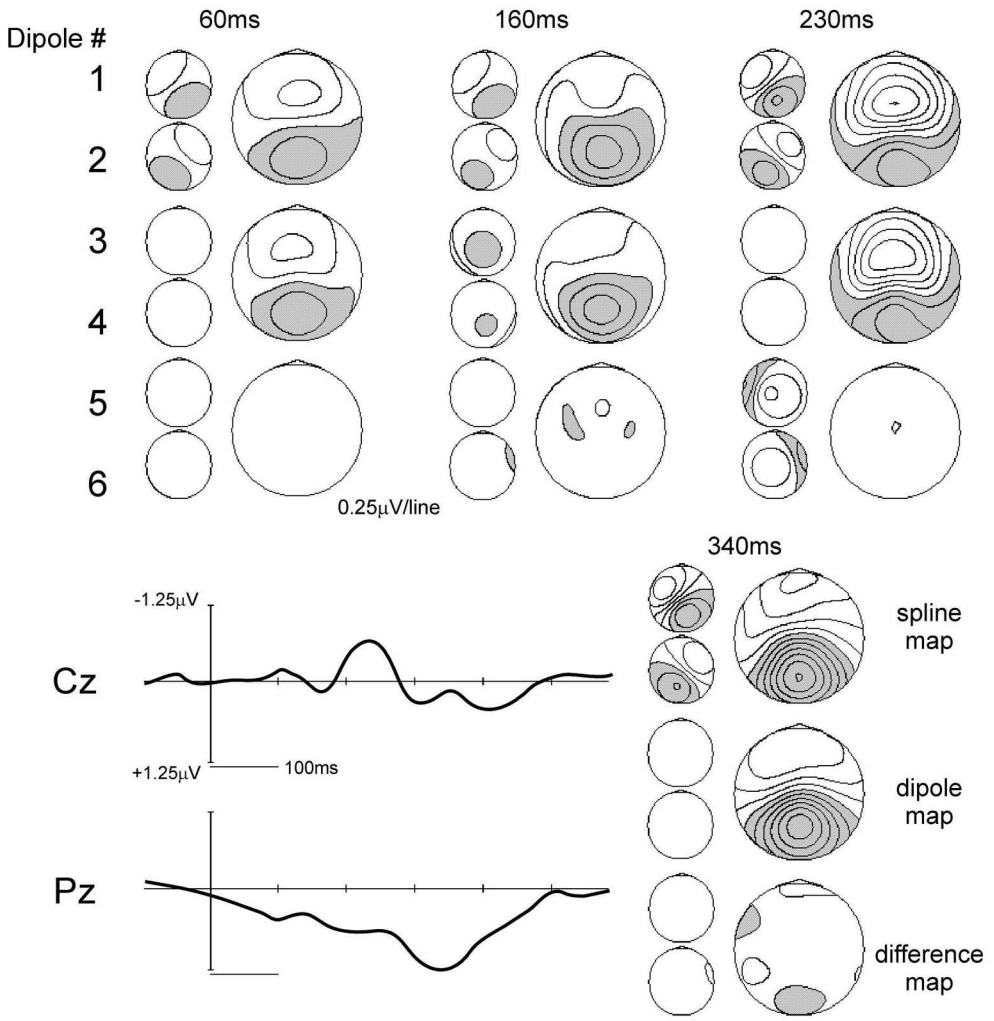

Fig. 6. Spline-interpolated voltage maps of the AV−(A+V) difference waveform at various time points. The six small maps represent the distribution of each dipoles' forward solution at that time point. From top to bottom, the three large maps represent the difference ERP distribution, the dipole model distribution and the difference between the first two, respectively. The waveforms at the lower left are the difference ERPs recorded at electrodes Cz, and Pz.

similar baselining problems could arise whether the sloping baseline was due to an anticipatory potential such as a CNV or to a long-latency slow potential elicited by the preceding stimuli. Superimposed upon these slow deflections in the AV−(A+V) difference wave were phasic waves that appeared to reflect true cross-modal interactions. These waves could be visualized more clearly in high-pass filtered waveforms that removed the anticipatory slow-potential shift. The first began at around 130 ms and peaked at 160–170 ms and could be modeled by a dipolar source in inferior occipital cortex. This interaction thus appears to take place in cortical areas of the occipito-temporal ventral stream. A similar auditory–visual interaction was observed at occipital sites by Giard and Peronnet [6] at 155–200 ms, which they interpreted as a modulation of the visual evoked N1 wave. This effect does indeed appear to represent an influence of auditory input on processing in a predominantly visual cortical area. The second major deflection indicative of cross-modal interaction peaked at 220–250 ms and could be accounted for by a dipole pair in anterior temporal peri-sylvian cortex. This effect might represent an interaction in auditory association cortex or in polymodal cortex of the superior temporal plane [3]. It should be noted, however, that the present procedures do not allow conclusions to be reached about the physiological nature of these cross-modal interactions that is, whether they are facilitatory or inhibitory or arise from cross-modal refractory interactions.

Some further comments on the rationale for analyzing the sources of the AV−(A+V) difference wave may be in order. This difference wave may have a non-zero value whenever the bimodal AV stimulus elicits a different amount or a different configuration of neural activity from the linear sum of the activities elicited by the unimodal A and V stimuli. This differential neural activity could result from an enhancement (or suppression) of activity within

one (or both) of the unimodal source configurations, or it could arise from a different source not activated by either unimodal stimulus alone. In any case, this differential neural activity due to cross-modal interaction is, in principle, localizable to a particular brain region or regions. Previous studies have made inferences about the localization of such interactive neural events on the basis of their associated scalp fields manifest in the AV−(A+V) difference wave [6,5]. In the present study, scalp topographical evidence was supplemented by dipole modeling in order to estimate the location(s) of the contributing source(s). Dipole modeling of difference wave topographies has also been used previously to estimate the sources of selective attention processes [7,20,2,12,14,13,11].

In conclusion, studies of cross-modal interactions in sensory processing need to take into account task-related neural activity that may be elicited following each stimulus but may not necessarily have anything to do with integration of information across different modalities. This caveat would seem to apply to studies using either electrophysiological or neuroimaging measures of brain activity, whenever the bimodal minus summed unimodal paradigm is applied. On the other hand, designs that do not assign any task-relevance to the unimodal and bimodal stimuli [5] would not be likely to encounter this problem. Further work is needed to establish clear criteria for distinguishing cross-modal interactions in sensory processing from more general task-related neural activity.

Acknowledgements

This study was supported by a grant from the National Institute for Mental Health (MH 25594). The authors thank Daniel R. Heraldez and Matthew M. Marlow for technical assistance.

References

[1] American Electroencephalographic Society, Guidelines for standard electrode position nomenclature, J. Clin. Neurophysiol 11 (1994) 111–113.
[2] L. Anllo-Vento, S.J. Luck, S.A. Hillyard, Spatio-temporal dynamics of attention to color: evidence from human electroencephalography, Hum. Brain Mapping 6 (1998) 216–238.
[3] G.A. Calvert, R. Campbell, M.J. Brammer, Evidence from functional magnetic resonance imaging of crossmodal binding in the human heteromodal cortex, Curr. Biol. 10 (2000) 649–657.
[4] V.P. Clark, S.A. Hillyard, Spatial selective attention affects early extrastriate but nor striate components of the visual evoked potential, J. Cogn. Neurosci. 8 (1996) 387–402.
[5] J.J. Foxe, I.A. Morocz, M.M. Murray, B.A. Higgins, D.C. Javitt, C.E. Schroeder, Multisensory auditory–somatosensory interactions in early cortical processing revealed by high-density electrical mapping, Cogn. Brain Res. 10 (2000) 77–83.
[6] M.-H. Giard, F. Peronnet, Auditory–visual integration during multimodal object recognition in humans: a behavioral and electrophysiological study, J. Cogn. Neurosci. 11 (1999) 473–490.
[7] H.-J. Heinze, G.R. Mangun, W. Burchert, H. Hinrichs, M. Scholz, T.F. Münte, A. Gös, M. Scherg, S. Johannes, H. Hundeshagen, M.S. Gazzaniga, S.A. Hillyard, Combined spatial and temporal imaging of brain activity during visual selective attention in humans, Nature 372 (1994) 543–546.
[8] S.A. Hillyard, The CNV and human behavior, in: W.C. McCallum, J.R. Knott (Eds.), Event-related Slow Potentials of the Brain: Their Relation To Behavior, Elsevier, Amsterdam, 1973, pp. 161–171.
[9] D.A. Jeffreys, J.G. Axford, Source locations of pattern-specific components of human visual evoked potentials. I: Components of striate origin, Exp. Brain Res. 16 (1972) 1–21.
[10] J.L. Kenemans, J.M.P. Baas, G.R. Mangun, M. Lijfijt, M.N. Verbaten, On the processing of spatial frequencies as revealed by evoked-potential source modeling, Clin. Neurophysiol. 111 (2000) 1113–1123.
[11] J.L. Kenemans, M. Lijffijt, G. Camfferman, M.N. Verbaten, Split-second sequential selective activation in human secondary visual cortex, J. Cogn. Neurosci. (in press).
[12] J.J. Lange, A.A. Wijers, L.J.M. Mulder, G. Mulder, ERP effects of spatial attention and display search with unilateral and bilateral stimulus displays, Biol. Psychol. 50 (1999) 203–233.
[13] G.R. Mangun, H. Hinrichs, M. Scholz, H.W. Mueller-Gaertner, H. Herzog, B.J. Krause, L. Tellman, L. Kemna, H.-J. Heinze, Integrating electrophysiology and neuroimaging of spatial selective attention to simple isolated visual stimuli, Vis. Res. 41 (2001) 1423–1435.
[14] A. Martinez, L. Anllo-Vento, M.I. Sereno, L.R. Frank, R.B. Buxton, D.J. Dubowitz, E.C. Wong, H. Hinrichs, H.-J. Heinze, S.A. Hillyard, Involvement of striate and extrastriate visual cortical areas in spatial attention, Nature Neurosci. 2 (1999) 364–369.
[15] J.J. McDonald, W.A. Teder-Sälejärvi, S.A. Hillyard, Involuntary orienting to sound improves visual perception, Nature 407 (2000) 906–908.
[16] M. Scherg, Fundamentals of dipole source potential analysis, in: F. Grandori, M. Hoke, G.L. Romani (Eds.), Auditory Evoked Magnetic Fields and Electric Potentials, Advances in Audiology, Vol. 5, Karger, Basel, 1990, pp. 40–69.
[17] B.E. Stein, M.A. Meredith, The Merging of the Senses, MIT Press, Cambridge, MA, 1993, pp. 123–156.
[18] W.G. Walter, R. Cooper, V.J. Aldridge, W.C. McCallum, A.L. Winter, Contingent negative variation, Nature 203 (1964) 380–384.
[19] R.B. Welsh, D.H. Warren, Intersensory interactions, in: K.R. Kaufman, J.P. Thomas (Eds.), Handbook of Perception and Human Performance, Sensory Processes and Perception, Vol. 1, Wiley, New York, 1986, pp. 1–36.
[20] A.A. Wijers, J.J. Lange, G. Mulder, L.J.M. Mulder, An ERP study of visual spatial attention and letter target detection for isoluminant and nonisoluminant stimuli, Psychophysiology 34 (1997) 553–565.