Preattentive detection of rare audiovisual feature conjunctions by the human brain as reflected by the mismatch negativity

Aarni Mustonen

Master's thesis
University of Jyväskylä
Department of Psychology
December 2007

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my supervisor Piia Astikainen for her guidance and valuable comments. I would also like to thank Lauri Viljanto for major and Kalervo Kela and Petri Kinnunen for minor technical assistance. Jukka Kaartinen’s patience merits to be praised as he repeatedly tolerated me entering his lab at an inappropriate moment during the data analysis phase of my study. I am also grateful to my brother Karri for his advice that helped me really start working. Finally, I want to acknowledge that Jenni’s love and support have encouraged me through my research process.

ABSTRACT

Preattentive detection of rare audiovisual feature conjunctions by the human brain as reflected by the mismatch negativity

Author: Aarni Mustonen
Supervisor: Piia Astikainen
Master's thesis in psychology
University of Jyväskylä
December 2007
39 pages

An influential view about feature binding, the feature integration theory (FIT), states that attention is needed in conjoining different features of a sensory object together. However, studies using the mismatch negativity (MMN), a component of the event-related potential (ERP) associated with automatic change detection after repeated sensory stimuli, have demonstrated that the MMN generating system can detect rare combinations of sensory features even though the features themselves are not rare. The purpose of the current study was to test whether unattended rare audiovisual stimulus combinations elicit an MMN. Event-related potentials (ERPs) were recorded from the scalp of ten female subjects while they were presented audiovisual oddball stimuli. Audiovisual stimuli consisted in visually presented bars with two alternating orientations paired with auditory tones with two alternating frequencies. Both bar orientations and both frequencies were presented at equal probability in both standard and deviant stimuli. However, in the standard stimuli, which were 90 per cent of the stimulus presentation trials, the higher frequency was combined with one orientation and the lower frequency, with the other orientation. In the remaining 10 per cent of the trials, the deviants, this rule was broken. During this audiovisual experiment, the subjects performed a somatosensory task, which prevented them attending the audiovisual stimuli. Differences in ERPs to standards and to deviants were obtained in three different time windows. At 90-110 ms after the stimulus onset, the difference in ERPs was negative at parietal and positive at frontal electrodes. At 160-180 ms period, the differential ERPs were negative at left hemisphere and at midline electrodes. At 330-350 ms time window, the potentials to deviants were more positive at one electrode, C4. This pattern of results is concluded to represent an audiovisual MMN, the MMN itself being either one of the two first differential peaks or some kind of double process consisting of both of these peaks. The implications of the results for the FIT are discussed and possible directions for future research are suggested.

Keywords: Audiovisual processing, Sensory memory, Preattentive processing, Feature integration, Event-related potentials (ERPs), Mismatch negativity (MMN)

TIIVISTELMÄ

Harvinaisten audiovisuaalisten piirreyhdistelmien tarkkaavaisuudesta riippumaton erottaminen ihmisaivoissa poikkeavuusnegatiivisuuden heijastamana

Tekijä: Aarni Mustonen
Ohjaaja: Piia Astikainen
Psykologian pro gradu –tutkielma
Jyväskylän yliopisto
Joulukuu 2007
39 sivua

Vaikutusvaltainen näkemys piirteiden yhteensitomisesta, piirreintegraatioteoria, esittää, että aistiobjektin eri piirteiden yhdistäminen vaatii tarkkaavaisuutta. Kuitenkin poikkeavuusnegatiivisuutta, toistuvien aistiärsykkeiden jälkeisen ärsykemuutoksen automaattiseen erottamiseen liittyvää aivojen herätevasteiden komponenttia käyttävät tutkimukset ovat osoittaneet, että poikkeavuusnegatiivisuuden synnyttävä järjestelmä kykenee erottamaan ärsykepiirteiden harvinaiset yhdistelmät, vaikka itse piirteet eivät ole harvinaisia. Tämän tutkimuksen tarkoitus oli testata, aiheuttavatko tarkkailemattomat harvinaiset audiovisuaaliset ärsykeyhdistelmät poikkeavuusnegatiivisuuden. Kymmenen naispuolisen koehenkilön päänahalta mitattiin herätevasteita kun heille esitettiin audiovisuaalisia oddball-ärsykkeitä. Audiovisuaaliset ärsykkeet koostuivat visuaalisesti esitetystä palkista, jolla oli kaksi eri asentoa, sekä äänistä, joita oli kahta eri korkeutta. Molemmat palkin asennot ja äänen korkeudet esiintyivät yhtä todennäköisesti sekä yleisissä standardiärsykkeissä että harvinaisissa devianttiärsykkeissä. Kuitenkin standardiärsykkeissä, joita oli 90 % kaikista ärsykkeistä, korkeampi ääni yhdistyi tiettyyn palkin asentoon ja matalampi ääni toiseen. Lopuissa 10 prosentissa ärsykkeistä, devianteissa, tämä sääntö rikkoutui. Tämän audiovisuaalisen kokeen aikana koehenkilöt suorittivat somatosensorista tehtävää, joka esti audiovisuaalisten ärsykkeiden tarkkailun. Standardeille ja devianteille saatiin eroavat herätevasteet kolmen aikaikkunan kohdalla. 90-110 ms ärsykkeen alun jälkeen erotusvaste oli negatiivinen parietaalisilla ja positiivinen frontaalisilla elektrodeilla. Aikavälillä 160-180 ms erotusvaste oli negatiivinen vasemman aivopuoliskon ja keskilinjan elektrodeilla. Aikavälillä 330-350 ms devianttien vaste oli positiivinen yhdellä elektrodilla, C4:llä. Näiden tulosten tulkitaan sisältävän poikkeavuusnegatiivisuuden niin, että itse poikkeavuusnegatiivisuus on joko jompikumpi kahdesta ensimmäisestä erotusvasteen huipusta tai jonkinlainen näistä molemmista koostuva kaksoisprosessi. Tulosten merkitystä piirreintegraatioteorian suhteen sekä jatkotutkimuksen mahdollisia suuntia käsitellään.

Avainsanat: Audiovisuaalinen prosessointi, Aistimuisti, Tarkkaavuudesta riippumaton prosessointi, Piirteiden integraatio, Herätevasteet, Poikkeavuusnegatiivisuus

CONTENTS

1. INTRODUCTION 1

1.1 The feature integration theory and the role of attention in visual search 1

1.2 The behavioral studies on feature integration in the auditory modality 4

1.3. Event-related potentials 8

1.3.1. The mismatch negativity 8

1.3.1.1. MMNs to auditory feature combinations 9

1.3.1.2. MMNs to visual feature combinations 13

1.3.1.3. MMNs to audiovisual feature combinations 13

1.4. The purpose of the present study 18

2. METHOD 19

2.1. Subjects 19

2.2. Stimuli and procedure 19

2.2.1. The audiovisual oddball experiment 19

2.2.1.1. The auditory stimulus components 20

2.2.1.2. The visual stimulus components 20

2.2.2. The somatosensory stimuli 20

2.2.3. The task 21

2.3. EEG recording 22

2.4. Data analysis 22

3. RESULTS 23

3.1. Behavioral task 23

3.2. ERP results 23

3.2.1. 90-110 ms poststimulus period 23

3.2.2. 160-180 ms poststimulus period 26

3.2.3. 330-350 ms poststimulus period 26

4. DISCUSSION 27

REFERENCES 33

1. INTRODUCTION

1.1 The feature integration theory and the role of attention in visual search

One of the key questions in neuroscience is the binding problem: how are different sensory stimulus features conjoined together to form a unitary percept. One central issue in feature binding is the role the attention plays in the process. A focal theory in this field of research is the feature integration theory (Treisman, 1986, 1988; Treisman & Gelade, 1980; Treisman & Schmidt, 1982), from this on shortened the FIT. The theory deals with object detection on the basis of different visual stimulus dimensions (such as color and form) that have certain values (for example, red or circular) called features. The FIT suggests that while single features are spotted automatically and in parallel, objects forming combinations of different features are scanned location by location with focused attention. This distinction was made on the basis of systematic differences in search times for single feature and conjunction targets in visual search tasks requiring the subjects to search for the presence of a predetermined target item in a matrix filled with distracting items (Treisman & Gelade, 1980). In a single feature search condition, the distractors lack the key feature present in the target item. In the conjunction search, instead, the features of the target item are also present in the distractors, and the target item is distinguished only by the presence of the combination of these features (for example, a red circle among red squares and blue circles). As the feature targets pop out from the distracting items, the search times for them are largely unaffected by increases in the number of the distractors. Search times for conjunction targets, instead, grow linearly in relation to the increases in the display size because each item displayed must be attended in its turn until the target is found. As evident from this pattern of search times, the conjunction targets are generally detected slower than the feature targets, a phenomenon called conjunction cost. Another finding predicted by the theory was the emergence of illusory conjunctions. Treisman and Schmidt (1982) demonstrated that attention is diverted with a distracting task or when the stimuli are presented for a very brief time, the features of one object can be erroneously perceived to belong to another object.

Treisman’s original formulation (Treisman, 1986) suggested that the preattentive feature detection takes place in specialized feature modules and the attentive feature binding accesses the representations of these features through a master map of locations. In more detail, the presence of individual features is initially coded by specialized feature modules into feature maps that indicate the presence of a feature without specifying where it is. These feature maps are linked to a master map of locations that shows where all feature boundaries are located but lacks specific information about which features are where. When features are bound together, attention focuses on a particular location on the master map and activates its links to the feature maps associated with the location. However, this model ran in trouble. Houck and Hoffman (1986) studied the McCollough effect (McCollough, 1965), a tendency to see colorless gratings as colored ones after having adapted to looking colored gratings. After adapting to green and black vertical columns and red and black horizontal rows, the subjects, when presented the same gratings in black and white, saw the vertical white columns as greenish and horizontal white rows as reddish (Houck & Hoffman, 1986). The authors found that the strength of the effect was unaffected by the state of the subjects’ attention. This meant that in the case of McCollough effect, combinations of color and orientation where not affected by the attention. With results like this in mind, Treisman (1988) reformulated the FIT, placing the master map of locations before the specialized feature modules. In this, later version of the theory, there are initial conjunctions present at the level of the master map of locations. However, these combinations cannot be interpreted in a form that is useful to the organism, until they are analyzed by specialized feature modules and then recombined under the spotlight of attention. This modification, named by Quinlan (2003) the Version 2 of the FIT, adjusted the protective belt of the theory to the new findings while maintaining the core assumptions about the automatic, preattentive and parallel feature detection and the attentive, serial conjunction detection.

The issues covered by the FIT remain controversial. As noted by Quinlan (2003) there have been many studies that have reported examples of conjunction detection in the absence of attention. Moreover, there are also examples that appear to indicate a role for attention in featural detection (Quinlan, 2003). The dichotomy between parallel, automatic feature detection and serial, attentive conjunction search is questioned by views that in general, both of the types of search occur in parallel, but the conjunction search is somehow more complex than the feature search and thus more prone to be distracted (Carrasco & McElree, 2001; Carrasco & Yeshurun, 1998; McElree & Carrasco, 1999). To give an example of the data supporting this alternative view, consider the study by Huang and Pashler (2005). They compared difficult search performance with a feature target, a conjunction target or a target spatial configuration in different distractor conditions. First, they compared the performance with a small, 8-item matrix to the performance with a larger, 16-item matrix. In all target conditions, the search times were significantly longer in the larger matrix condition, a finding that might implicate a role for attention in the search performance. The authors, however, went further and investigated the performance with the 16-item matrix in two different display conditions. In the simultaneous condition, all items of the matrix were presented simultaneously. In the successive condition, one half of the matrix was presented first and the other half after it. If the search performance had been faster in the successive condition than in the simultaneous condition, it would have meant that being able to attend just one part of the display enhanced the performance, which could have been interpreted as a sign of attentional capacity limiting the search performance in the simultaneous condition. However, the simultaneous and successive conditions did not differ significantly in either feature or conjunction searches, indicating that the attentional capacity does not limit search performance for feature or conjunction targets. The authors argue that it is statistical decisional noise and possibly the amount of necessary eye movements, not attentional capacity limits that made their subjects' search performance slower with the increase of the number of the distractors. Although Huang and Pashler do not dwell on the issue, it seems possible that the typically different effects of the display size on the search times for feature and conjunction targets reflect their differences in the statistical decisional noise as the amount of eye movements needed in the two target conditions does not differ. To conclude, no consensus exists over the question whether the attention has a special role in visual feature integration and what the role might be.

Although originally a theory of visual cognition, the FIT makes claims about the nature of sensory processing that, at least possibly, might also be generalized to processing outside the visual modality. The behavioral studies addressing the applicability of the FIT to auditory processing are considered in the next section.

1.2 The behavioral studies on feature integration in the auditory modality

The generalizability of the FIT into the domain of audition has received little attention compared to the vast literature on visual feature binding. The behavioral studies on the issue can be divided into the studies of illusory conjunctions and into reaction time studies addressing the effects of different auditory dimensions and their combinations on the responses of the subject. Most of the studies in both groups have used sequentially organized sounds in contrast to the spatial distribution of objects in visual search matrices. This is a consequence of the fact that the simultaneous presentation of such a large number of auditory objects as the number of objects in a visual search task would result in peripheral masking at the cochlea (Woods, Alain, & Ogawa, 1998). Let us first consider the illusory conjunction studies and then move to the reaction time studies.

Illusory conjunctions in the auditory modality have been reported not only with stimuli presented in a serial sequence (Thompson, Hall, & Pressing, 2001) but also with spatially distributed sounds (Hall, Pastore, Acker, & Huang, 2000; Takegata, Brattico, Tervaniemi, Varyagina, Näätänen, & Winkler, 2005). Hall et al. presented their subjects arrays of two or four sounds. The subjects were instructed to indicate whether cued conjunctions of pitch and timbre were present in the array. The subjects frequently indicated the presence of the cued conjunction when its pitch was present in one sound and its timbre in another sound of the array. This indicates a failure in combining the two features. These results where later replicated with arrays of two sounds by Takegata et al. (2005). The study of Thompson et al. (2001) appeared to confirm that illusory conjunctions also occur when stimuli are presented serially. The authors presented their subjects target sequences of two or seven tones and after the sequences, a probe tone. The subjects were instructed to indicate whether the probe tone matched one of the target tones in both pitch and duration. The subjects made few errors when only one of the two defining features of the probe tone were present in a target sound, but when the duration of the probe was present in one tone and the pitch in another, the subjects made many errors. However, Jamieson, Thompson, Cuddy, and Mewhort (2003) came up with an alternative explanation of these results. Basing their argument on global recognition memory theory (Clark & Gronlund, 1996; Humphreys, Pike, Bain, & Tehan, 1989) and supporting their argument with mathematical simulations, the authors suggest that the similarity structure between the probe and the target tones used by Thompson et al. predicts a pattern of results similar to the one reported by those authors. This practically means that the results of Thompson et al. (2001) did not necessarily reflect a failed feature binding process. Jamieson et al. also speculate that similarity between the probe and the stimulus array might have influence on the results in other studies reporting auditory illusory conjunctions.

Illusory conjunctions are an important phenomenon but the key issue regarding the role of attention in auditory feature binding and the generalizability of the FIT into auditory processing is about conjunction cost: are conjunctions never detected and processed faster than the analysis of their slowest feature? In this line of research, the reaction times are of interest. A conjunction benefit, that is, a feature conjunction receiving a quicker reaction than some of its constituent features, would practically mean that auditory features do not have to be conjoined by focused attention after their features are first analyzed. This would mean that the FIT does not apply to auditory processing. This was exactly what was reported by Woods et al. (1998). They instructed their subjects to respond to the presence of a prespecified sound feature (a certain frequency or a certain location) or of a prespecified conjunction of these features. Under a high-rate serial presentation of stimuli, their subjects reacted fastest to the targets defined by their frequency. What is most important, however, is the fact that the conjunction targets received faster reactions that the targets identified solely by their location.

Woods, Alain, Diaz, Rhodes, and Ogawa (2001) went on to further elaborate the results of Woods et al. (1998). They replicated the finding of conjunction benefit, this time with combinations of duration, frequency and location that where processed faster than conjunctions of duration and location. The authors proposed that in auditory selective attention tasks the frequency plays an organizing role analogous to the role of spatial position in visual attention. These results and views were, however, criticized by Dyson and Quinlan (2003). They replicated the faster processing of conjunctions, this time with conjunctions of auditory phonemes and their locations. These conjunctions where processed faster than targets defined by either of these features. Nonetheless, the authors argued that this kind of tasks do not necessarily require feature binding but the result pattern could be explained in terms of feature coactivation (Miller, 1982; Mordkoff & Yantis, 1991; Mordkoff, Yantis, & Egeth, 1990). The feature coactivation models suggest that a decision about a presence of a stimulus is based on a mechanism that pools information from individual feature analyzers. Thus, the conjunction benefits could have resulted from the fact that conjunction targets activate two feature detectors and thus the total activation builds up faster than when there is only one active feature detector. This could lead to faster detection of conjunctions even if the constituent features are not bound together. Dyson and Quinlan went on to experimentally study what happened when a so-called condensation task was used in the conjunction condition. In this task, participants were forced to make judgments about the presence or absence of the target conjunctions themselves, as all their constituent features where also present in nontarget conjunctions. The results indicated that the conjunction benefit had disappeared, that is, the conjunctions where detected slower than feature targets, as predicted by the FIT.

If the conjunction benefits reported in the literature are due to feature coactivation processes, does this mean that the auditory processing occurs exactly in a manner described by the FIT? Not necessarily. Mondor, Zatorre, and Terrio (1998) reported findings indicating that their subjects could not completely ignore either the frequency or the location of a tone. Even when the target sounds could in principle be distinguished based on just one of these dimensions, variance on the irrelevant dimension slowed down the detection of the target sounds. This appears to indicate a holistic processing of all sound features together. This picture was later made clearer by Dyson and Quinlan (2004). They presented their subjects same-different matching tasks first with long (750 ms) and then with short (100 ms) inter-stimulus intervals (ISIs) between a first, target tone and a second tone that the subjects matched to the first. In both ISI conditions, there was a frequency matching task and a location matching task. In the frequency task the participants indicated whether the frequency of the second sound matched the frequency of the first tone. The changes in the location between the two tones were irrelevant in the frequency task. In the location task, the relevant, matched dimension was the location, while the frequency was irrelevant. The results in both ISI conditions show that the stimuli that were exactly similar than the one before them received fastest reactions, while changes in the irrelevant dimensions slowed down the matching process. This suggests a holistic perception with both location and frequency information as an integral part of the perception. However, there were also signs of separate analytic processing of both these features. In the short ISI condition, when there was less time for coding the first stimulus before the presentation of the second, matching task stimulus, the relevant changes in the location condition received faster responses than the relevant changes in the frequency condition. This clearly suggests that the features where processed individually. Dyson and Quinlan discuss their findings in terms of a dual processes account of processing in which the new stimuli first receive a holistic check, which tells the auditory system whether the sound is completely similar to the previous one. After the initial holistic processing, they argue, the different features are analyzed by separate feature modules. Dyson and Quinlan conclude that this pattern fits the Version 2 of the FIT with master map located before the feature detectors (Treisman, 1988). However, there is actually nothing in their data that indicates that the holistic check is performed before the analysis of individual features. The fact that the short ISI revealed the differences in the processing of the two features does not mean that these processes occur later than the holistic stimulus perception. It was the encoding of the first tone that was impeded by the short ISI, not the analysis of the features of the second tone in relation to the features of the first one. Thus, in fact, the holistic perception and individual feature analysis could have taken place, for example, simultaneously in different anatomical locations in the brain.

As controversies remain in the interpretation of behavioral results on auditory feature binding, the focus of auditory feature binding research is shifting toward electroencephalography. Since the possibilities of spatial stimulus distribution of auditory stimuli are very limited, the need for tools for studying serially presented stimuli is great. The electroencephalographic recordings hold much promise in feature binding research because they enable the study of processing of unattended stimuli, including feature combinations. Let us begin with the introduction of the event-related potentials in general and, after that, the mismatch negativity component, which has a crucial role in unraveling preconscious sensory processing (Näätänen, Tervaniemi, Sussman, Paavilainen, & Winkler, 2001).

1.3. Event-related potentials

Event-related potentials (ERPs) are electric brain responses elicited by certain events or stimuli presented in a serial sequence. The ERPs are obtained from the electroencephalogram (EEG), which is a non-invasive means of recording the brain's electrical activity through the meninges, the skull and the skin. The ERPs also have a magnetic equivalent, the event-related fields, that are obtained using the magnetoencephalography (MEG).

The raw EEG signal is converted into ERPs by averaging the signals received at the time of each stimulus presentation (trial), a process that increases the amount of signal in relation to the amount of noise. The ERPs have a high temporal resolution, giving a millisecond-by-millisecond picture of the cognitive process associated with the stimulus or event. The attempts for spatial localization of the sources of the ERPs in the brain must, however, deal with so-called inverse problem, because there is no unequivocal means to determine the number of brain generators behind the signal measured on the scalp (Hari, 1994).

1.3.1. The mismatch negativity

The mismatch negativity (MMN) is an event-related potential introduced by Näätänen, Gaillard, and Mäntysalo (1978). It has been studied mostly in the auditory domain, but also the visual form of MMN (for a review, see Pazo-Alvarez, Cadaveira, & Amenedo, 2003) is becoming a widely studied phenomenon. Peaking between 100 and 250 ms after the stimulus onset, the MMN is a negative component in the EEG curve, elicited by an automatic detection of a disruption of regularity in the stimulus stream. The MMN is traditionally studied in an auditory oddball paradigm in which often occurring standard stimuli are occasionally replaced by rarely occurring deviant stimuli differing from the standards in some physical feature such as, for example, the sound frequency.

The MMN results from a comparison process between an incoming deviant stimulus and the representation of the standard stimulus in the sensory memory (Näätänen, Jacobsen, & Winkler, 2005). The auditory MMN has its maximal amplitude at the frontocentral EEG electrodes, showing a reversed, positive polarity at the mastoid electrodes in the lower temporal area. The auditory MMN appears to be produced by two different neural generators (Shalgi & Deouell, 2007). According to the prevailing theory, the temporal generator is involved in generating sensory memory traces and comparing incoming stimuli with these traces; the frontal generator is responsible for triggering an involuntary attention switch toward the detected change (Näätänen & Michie, 1979).

The MMN can give an objective measure of the brain's ability to distinguish between different auditory stimuli such as phonemes (Näätänen, 2001). Being independent of attention, the MMN is a suitable tool for studying a wide range of patient groups from developmental and neurological to psychiatric disorders (Näätänen, 2003). The MMN has traditionally been used to study the detection of stimuli deviating from the standards in one stimulus feature. However, the independence of attention of the MMN makes the MMN also an ideal tool for investigating whether multiple unattended stimulus features are bound together to form unitary percepts in the brain. It is relatively simple to design an ERP experiment in which the deviant stimuli do not contain rare stimulus features but rare combinations of the same features that are also present in the more frequent, standard combinations. This kind of studies are reviewed in the next sections.

1.3.1.1. MMNs to auditory feature combinations

Auditory feature binging has been studied with the MMN for a decade. The first study of this kind was reported by Gomes, Bernstein, Ritter, Vaughan, and Miller (1997) who found rare, deviant combinations of sound frequency and intensity to elicit an MMN. After that, MMNs have also been reported to deviant combinations of frequency and location (Sussman, Gomes, Nousak, Ritter, & Vaughan, 1998), frequency and timbre (Takegata et al., 2005), frequency and duration (Paavilainen, Arajärvi, & Takegata, 2007), and frequency, intensity and duration (Ruusuvirta & Huotilainen, 2004).

Two published studies have focused on the question about whether the detection of deviant feature conjunctions occurs independently from the detection of deviant single features. Takegata, Paavilainen, Näätänen, and Winkler (1999) compared the amplitudes of the MMNs to location deviants and to deviant combinations of frequency and intensity to the MMN amplitude to so called double deviants deviating from the standards simultaneously in the location and the frequency-intensity combination. The MMN amplitude to double deviants was found to equal the sum of the MMN amplitude to combination deviants and the one to location deviants, indicating an independent processing of deviating feature combinations from deviating single features. Accordingly, Takegata, Huotilainen, Rinne, Näätänen, and Winkler (2001) used equivalent current dipoles (ECD) of magnetic mismatch negativity (MMNm) and found out that deviant combinations of frequency and location are processed at least partially by different neuronal populations than stimuli deviating from standards in either one of these features.

The ability to distinguish between rare and often occurring feature combinations is not limited to adult humans. Ruusuvirta Huotilainen, Fellman, and Näätänen (2003) discovered a positive, MMN-like differential potential to rare combinations of sound frequency and intensity in sleeping newborn infants, suggesting that the auditory feature binding occurs in the human brain already in the first days after birth. Indeed, not only the human brain binds unattended sound features together, as demonstrated by Astikainen, Ruusuvirta, Wikgren, and Penttonen (2006) who found the rare conjunctions of frequency and intensity to cause a positive, MMN-like differential potential in the epidural surface of urethane-anesthetized rats.

The MMN generating system of the brain is capable of extracting complex rules from sound sequences (for a review, see Näätänen et al., 2001). For example, Saarinen, Paavilainen, Schöger, Tervaniemi, and Näätänen (1992) reported an MMN to sound pairs with rare direction of frequency change in an experiment in which all the sound pairs constantly varied in frequency. Recently, the experiment was replicated with urethane-anesthetized rats, resulting in higher absolute ERP amplitudes to the tone pairs with deviant direction of frequency change than to the standard pairs (Ruusuvirta, Koivisto, Wikgren, & Astikainen, 2007). The rule extracting ability of the brain raises an interesting question: are deviant feature combinations detected because the brain makes a distinct memory representation of each feature combination and compares the probabilities of their occurrence? Or does the brain only extract the rules governing how the features are usually combined and detect the stimuli breaking these rules? Of course it is also possible that both of these mechanisms operate simultaneously. Let us start considering the empirical evidence with a study by Paavilainen, Simola, Jaramillo, Näätänen, and Winkler (2001). They found an MMN elicited by deviant combinations of frequency and intensity in an experiment in which there were eight standard and eight deviant combinations that could be distinguished from each other by applying a simple rule (the higher the frequency, the higher the intensity should be). In another condition of the experiment, the standards where eight combinations formed on a completely random basis. The deviants where also eight random combinations of frequency and intensity. In this condition, the standards and the deviants could not be distinguished on the basis of a rule. No MMN was elicited, which suggests that rules are necessary for automatic detection of feature conjunctions at least if the number of the conjunctions is very high. Ruusuvirta and Huotilainen (2004), on the other hand, found an MMN elicited by the deviant combinations in an experiment of six standard and six deviant combinations of frequency, intensity and duration that could not be distinguished on a basis of a few simple rules. Furthermore, when the same stimuli were presented to the newborn infants already mentioned in the context of another study (Ruusuvirta et al., 2003), the six deviant feature combinations evoked a positive, MMN-like differential potential (Ruusuvirta, Huotilainen, Fellman, & Näätänen, 2004). The most straightforward interpretation of this pattern of results seems to be that both rules and separate memory representations of each combination are used in the automatic detection of feature conjunctions.

A few recent studies have focused on the relationship between the attention and the auditory feature binding indexed by the MMN. Winkler, Czigler, Sussman, Horváth, and Balázs (2005) systematically manipulated the attentive demands of the task their subjects performed during the presentation of combinations of frequency and sound location and found the MMN elicited to be independent of attention. Similar results were also obtained by Takegata et al. (2005). As the other MMN studies of feature binding differ from visual search tasks, that form the basis of the FIT, because in ERP studies the stimuli are presented in a serial sequence, Takegata et al. presented their subjects with concurrent spatially distributed sounds forming conjunctions of pitch and timbre. This made their experimental design resemble the visual search task a little more than the usual, serial stimulus presentation. During the stimulus presentation, the subjects performed a visual working memory task. The visual task difficulty did not affect the amplitude of the MMN evoked by deviant conjunctions. In a separate, behavioral task using same stimuli as the ERP experiment, the subjects, while performing the visual task, were presented a probe sound and after it, an array of two concurrent sounds. The subjects were instructed to indicate whether the probe tone appeared among the array of two sounds. As it was already briefly mentioned earlier, illusory conjunctions occurred relatively frequently when a pattern of the target pitch appearing in one sound and the target timbre in the other sound was mistaken for the appearance of the target sound. Thus, the preattentive MMN generating system appeared to manage to integrate the two sound features but this information was not available to conscious perception. Moreover, Paavilainen et al. (2007) found that the human brain is, in absence of attention, capable of combining not only the features of one sound but also processing a predictive relationship between the features on one dimension of the first sound and the features on another dimension of the next sound. When the duration of the previous sound predicted the frequency of the next sound, the sounds violating this rule elicited an MMN. Furthermore, the subjects did not manage to figure out the rule between the duration and the frequency, and even when the rule was explained, the subjects had difficulty consciously detecting the deviant stimuli. Together these results form a coherent picture: sound features seem to be bound together preattentively.

In conclusion, the evidence from MMN studies for preattentive auditory binding is strong. This preattentive feature binding occurs with a wide range of combinations of all different sound dimensions: frequency, intensity, location, timbre and duration. The feature combinations are processed in both adult and infant human brains and even in anesthetized rat brain. Attentive manipulations do not affect this automatic binding process in adult humans, and the MMN generating system seems to be capable of greater accuracy in making sense of the feature combinations than the conscious performance of the subjects. These findings clearly contradict the ideas of the original FIT that accepts no idea of preattentive feature binding of any kind. There also seems to be a conflict between these MMN results and the Version 2 of the FIT. This Version 2 (Treisman, 1988) suggests that different stimulus dimensions are initially conjoined at the master map of locations before being analyzed by different feature modules and then recombined with the spotlight of spatial attention. However, the preattentive feature binding indexed by the MMN requires the registration of the probabilities of occurrence of different feature combinations. This is a process that requires formation of organized transient memory traces and seems to fit the notion of primitive intelligence in the auditory cortex (Näätänen et al., 2001). This seems hard to negotiate with the notion of initial conjunction formation of the version 2 of the FIT that, although described very briefly, appears to be a description of a very passive detection process with no analysis of any kind being performed before the sensory signal enters the individual feature modules.

1.3.1.2. MMNs to visual feature combinations

In the visual domain, only one ERP study of unattended feature conjunctions has been reported. Winkler et al. (2005) studied the detection of rare combinations of crating orientation and color. The rare combinations elicited a similar MMN no matter whether the subjects were instructed to attend the combinations or the occasional changes of a fixation cross at the center of the visual field. This suggests that the visual feature conjunctions possibly receive a similar analysis by the MMN generating system as the auditory conjunctions do. However, more research is needed because it is risky to make conclusions on the basis of only one study. Additional research would be important because the visual modality is the sensory modality the FIT makes specific claims about.

1.3.1.3. MMNs to audiovisual feature combinations

Before proceeding to studies of MMN to audiovisual feature combinations it is reasonable to consider the different methods used for controlling the subjects' attention. In auditory MMN experiments it is common that the subjects are instructed to focus on a visual task (e.g. Takegata et al. 2005), read materials of their choice (e.g. Sussman et al. 1998) or watch a silent movie (e.g. Ruusuvirta & Huotilainen, 2004). In visual MMN experiments, on the other hand, the subjects usually must keep their eyes fixated on the screen where the visual stimuli are presented. This is usually achieved with visual stimuli belonging to a behavioral task, appearing at the center of a display, while the stimuli of experimental interest take place at other locations of the visual field (Czigler, Balázs, & Winkler, 2002; Czigler, Weisz, & Winkler, 2006; Pazo-Alvarez, Amenedo, & Cadaveira, 2004). However, an experimental design more analogous to the auditory MMN would instead of a visual task make the subjects attend an auditory task while maintaining the necessary eye fixation (Astikainen, Ruusuvirta, Wikgren, & Korhonen, 2004).

Also in audiovisual MMN studies the subjects are typically instructed to keep their eyes fixated on the display screen. The instructions about what the subjects should attend to, vary from one experiment to another. For making conclusions about binding of unattended audiovisual features, the most valid data comes from experiments in which the subjects are attending a task outside the visual and auditory modalities, such as a tactile task. Unfortunately, only three such (Colin, Radeau, Soquet, Dachy, & Deltenre, 2002a; Colin, Radeau, Soquet, & Deltenre, 2004; Colin, Radeau, Soquet, Demolin, Colin, & Deltenre, 2002b) audiovisual MMN studies have been reported. That is why results from studies with other kind of attentional manipulations need to be considered in this study.

In many audiovisual experiments, the experimental designs have included an instruction for the subjects to attend a visual task unrelated to the visual stimuli of experimental interest. In these kind of designs the amount of attention that leaks into processing the experimental stimuli is probably affected by the degree of similarity between the experimental stimuli and the attentional target stimuli, more similarity between them meaning more attention directed to the experimental stimuli. This view is in line with the attentional engagement theory (Duncan & Humphreys, 1989) which argues that in competition for the access into the visual short term memory, a stimulus gains weight to the extent that it matches the internal template of the information needed in the current behavior. Thus, it can be argued that an audiovisual experiment with attended target stimuli very different from the visual stimuli of experimental interest, can give us information about unattended binding of auditory and visual features, whereas a design with target stimuli that are very similar to the visual stimuli of experimental interest, can mean that the data received results from the integration of attended visual stimuli with unattended auditory stimuli.

In some audiovisual studies the subjects are given the instruction to attend the experimental stimuli either in one of the two sensory modalities or in both of them. Even these kind of manipulations can give information about preattentive feature binding if, and only if, the subjects are attending to one of the two modalities in which there is only one type of stimuli, and the changes in the stimuli of the unattended modality affect the processing of the attended, unchanging stimuli. In other kinds of designs with attended stimuli of experimental interest, no information about preattentive audiovisual feature binding is obtained. Having considered these methodological issues, let us proceed to the reported audiovisual MMN results.

An audiovisual MMN was first reported in the context of so-called McGurk-illusion. The McGurk illusion (McGurk & McDonald, 1976) can be observed when an incongruence takes place between an auditorily presented phoneme and another, visually articulated phoneme. Thus, for example, a combined presentation of an auditory bilabial consonant, such as /b/ in the syllable /ba/, and a visually articulated velar consonant, such as /g/ in the syllable /ga/, can lead to an illusory perception of an alveolar consonant, in this case /d/ in the syllable /da/.

The first study of the MMN to McGurk illusion was reported by Sams, Aulanko, Hämäläinen, Hari, Lounasmaa, Lu, and Simola (1991). Their subjects were instructed to count the number of auditory stimulus presentations. The auditory stimulus component presented was always the syllable /pa/. A simultaneous visual presentation of the articulation of either the syllable /pa/ or the syllable /ka/ led to audiovisual combinations in which the auditory and visual part were either concordant or discordant. MEG was used to measure the left brain hemisphere activity in the subjects during the experiment. The discordant stimuli evoked a negative differential magnetic response from the concordant ones, probably indicating a source in the supratemporal auditory cortex, similarly to the source of the auditory MMNm.

Colin et al. (2002b) reported an electric MMN evoked by deviant McGurk percepts from combinations of an auditory syllable /gi/ and a visual syllable /bi/. Importantly, their subjects were attending a tactile discrimination task during the

ERP recording. Two years later, the same group (Colin et al. 2004) tried to generalize their findings to voiceless consonants using syllables /pi/ and /ki/. They calculated the MMN by subtracting the ERPs to the deviant, incongruent audiovisual syllables that were presented alone, from the ERPs to deviants presented in an oddball paradigm with the standard, congruent stimuli. One of the two incongruent audiovisual syllables elicited an MMN and the other did not. The authors explained the failure to obtain an MMN to the other incongruent pair as a result of the visual articulatory movements used starting too early before the auditory syllable. The authors also warned that the N2b waveform, which is elicited when the stimuli were attended, could explain the differential potential obtained. However, as the subjects were attending a tactile task while keeping their eyes on the screen and as there was no P3 waveform that is also associated with attention and occurs often with the N2b, the N2b explanation doesn't seem probable. A more serious concern for Colin et al. (2004) was the possibility of refractoriness, that is, differences in habituation to stimuli, explaining their results. The deviants in the deviants alone –condition were presented more frequently than the deviants in the oddball paradigm including the standard stimuli. This means that the alone presented deviants were probably met by more habituation of the auditory system than the deviants in the oddball situation. The difference in the habituation level between the two conditions compared with ERPs might have meant that the differential potential obtained resulted from refractoriness, not from a memory process. Thus, the generalizability of the MMN to McGurk percepts of voiceless consonants remains uncertain.

Two other studies have also reported an MMN (Saint-Amour, De Sanctis, Molholm, Ritter, & Foxe, 2007) and a magnetic mismatch field (Möttönen, Krause, Tiippana, & Sams, 2002) elicited by illusory McGurk percepts. However, in Saint-Amour et al. (2007) the subjects were attending the visual syllables presented and in Möttönen et al. (2002), the audiovisual stimulus combinations. Thus, neither of these two results was obtained in absence of attention and they are of little significance for the FIT.

Another form of audiovisual illusion, the ventriloquist illusion, a tendency to ignore or underestimate the spatial separation between the sources of synchronous auditory and visual signal, was used by Colin et al. (2002a) to demonstrate a suppression of the auditory MMN. In an auditory condition, deviant sounds emerging from a location deviating 20° from the source of the standard sounds elicited an MMN. However, when in an audiovisual condition the same sounds were presented together with a visual signal that was always appearing in one unchanging location, all the sounds were perceived as coming from the same source and no MMN to the deviating sound location was elicited. As in other studies by the same group, the subjects were attending a tactile discrimination task during the experiment. Stekelenburg, Vroomen, and de Gelder (2004), instead of demonstrating an MMN suppression, reported an MMN elicited by an illusory sound shift. The subjects, attending to visual task unrelated to the experimental stimuli, were presented sounds always originating from the same location and, simultaneously, a light flashing in a usual standard and occasionally in a rare, deviant location. The location of the light induced a perception of the origin of the sound. The sounds associated with the light flash in the deviant location elicited a MMN very similar to one to real sound location shifts. The MMN obtained was truly an audiovisual one, as was shown when in the control condition, the visual stimuli, presented with no accompanying sounds, elicited no visual MMN.

McGurk and ventriloquist effects are both audiovisual illusions. Reports of MMN in the context of these phenomena raise the question: are audiovisual illusions a special case or can the MMN be elicited by audiovisual stimulation in the absence of illusory perceptions? This question was addressed by Besle, Fort, and Giard (2005). They presented their subjects audiovisual stimuli, of which the standard ones consisted in a deformation of a circle into an ellipse in the horizontal direction, paired with an auditory sound shift from 500 to 540 Hz. There where three types of deviant stimuli. The visual deviant paired the standard sound shift with a circle deformation in the vertical direction. The auditory deviant, in turn, paired the standard visual stimulus component with a rare sound shift from 500 to 600 Hz. The double deviants consisted in the rare forms of both the auditory and the visual stimulus components. The subjects' attention was directed to an unrelated visual task. The amplitude of the MMN to the double deviants did not equal the sum of the MMN amplitudes to the visual deviants and to the auditory deviants. Furthermore, the amplitude of the MMN to the visual deviants in the audiovisual experiment differed from the amplitude of the visual MMN obtained in a separate, completely visual control condition. These findings suggest that the auditory and visual stimulus features were stored as at least partly unitary memory representation.

In another audiovisual study not related to illusions, Rahne, Böckmann, von Specht, and Sussman (2007) demonstrated that while their subjects performed an unrelated visual task, visual cues determined whether two auditory streams were perceived as a unitary stream or as two streams. Only when the streams were segregated into two, the disruptions of regularity in the frequencies of the auditory stimuli in the lower stream could be detected and evoked an MMN.

Other studies on audiovisual MMN have used different attentional instructions making the subjects attend the visual stimuli (de Gelder, Böcker, Tuomainen, Hensen, & Vroomen, 1999; Surakka, Tenhunen-Eskelinen, Hietanen, & Sams, 1998; Ullsperger, Erdmann, Freude, & Dehoff, 2006) or even the combinations of the visual and auditory stimuli (Widmann, Kujala, Tervaniemi, Kujala, & Schröger, 2004). All these studies reported that the auditory MMN was either elicited or modulated by the visual information, but as these MMNs were not obtained in the absence of attention, they are of little significance for the FIT.

1.4. The purpose of the present study

The audiovisual MMN data appears to support the hypothesis that audiovisual feature conjunctions are processed preattentively by the MMN generating system in a similar manner as the auditory feature conjunctions are bound. For reasons similar to the ones discussed in the context of preattentive auditory feature binding, this would suggest that the FIT does not apply to audiovisual integration. However, as reviewing the studies presented above indicates, preattentive audiovisual feature binding has been mainly been studied with experiments designed for research problems tangential to the focus of this thesis. Most of audiovisual MMN studies have been conducted in the context of McGurk illusion, in which there are already long term memory representations about the visual and auditory properties of the phonemes before the experiment starts. Outside McGurk and ventriloquist illusions there have been very few studies published on the issue. None of these studies have used an experimental design directing the attention of the subjects toward a somatosensory or other third sensory modality task. Furthermore, no audiovisual MMN study has used a design with one auditory and one visual stimulus dimension, which both have two forms that have equal probability of occurrence and make four different audiovisual combinations, of which the two standard combinations are distinguished from the deviant combinations only by their probability of occurrence. This should cause brain make representations about the probabilities of the co-occurrence of the visual and auditory features. No effects of refractoriness, that is, more habituation of participants' afferent auditory pathways to some physical features of the stimuli than to other ones, can affect the ERPs whatsoever, as both the sounds and both the images have the same probability of occurrence. This study attempts to make a rigorous test for the existence of MMN to rare audiovisual feature combinations, which would indicate that auditory and visual features are integrated in the absence of attention.

2. METHOD

2.1. Subjects

The participants were ten female university students who were unaware of the purpose of the study and volunteered to participate. The subjects ranged in age from 19 years and 1 month to 21 years and 7 months, the mean age being 20 years and 3 months. All subjects had self-reported normal vision (corrected if necessary) and hearing and none of hem had a history of neurological diseases. All subjects gave written informed consent before the experiment. During recordings, the subjects were seated in a dimly lit room.

2.2. Stimuli and procedure

2.2.1. The audiovisual oddball experiment

The ERP-evoking stimuli were four different types of audiovisual stimulus combinations of 100 ms in duration presented at 505 ms stimulus-onset-asynchrony (SOA) in an oddball experiment of 1500 trials. Of the four stimulus combinations, two equiprobable standards occurred at the probability of 0.45 and two equiprobable deviants, at the probability of 0.05. The two kinds of tones and the two kinds of visually presented bars formed biconditional combination patterns, which could not be separated from each other by relying on visual or auditive information alone. For the first half of the participants, one standard was a 1000 Hz tone with a bar tilted clockwise and the other standard, a 1500 Hz tone with a bar tilted counterclockwise. For these subjects, one deviant was a 1000 Hz tone with a bar tilted to the right and the other deviant, a 1500 Hz tone with a bar tilted to the left. For the second half of the participants, the stimulus assignment for the standards and the deviants was reversed. The stimulus presentation was controlled using Psychology Software Tools E-Prime software.

2.2.1.1. The auditory stimulus components

The auditory components of the audiovisual combinations were 1000 Hz and 1500 Hz sinusoidal tones of 100 milliseconds in duration (including 10 ms rise and fall times). The tones were presented from a loudspeaker located above and slightly to the left of the head of the participants at the intensity of 65 dB.

2.2.1.2. The visual stimulus components

The visual components of the audiovisual combinations were black bars presented on a white background on an Eizo Flexscan F53 computer screen that was located 1 m in front of the subjects and covered 17° x 13° of their visual field. One kind of bars was tilted 16° clockwise from the vertical position and other kind, 16° counterclockwise. The bars were 24,5 cm long and 6 cm wide, appearing on a vertical area of 13° and on a horizontal area of 7° of the visual field of the participants.

2.2.2. The somatosensory stimuli

Before the experiment, two hoses were attached in the right hand of the subjects. One hose was attached between the index and middle fingers, the other between the ring and little fingers. Puffs of air of duration of 100 ms were delivered through these hoses during the experiment. The presentation of these air puffs was controlled using the E-PRIME program in a different computer than the one controlling audiovisual ERP stimuli. The occurrence of the air puffs was timed in a completely random way in relation to the audiovisual stimuli. On a pseudorandom basis, one out of 11 air puffs, the target puffs, took place in the hose located between the ring and little fingers of the participants. There were never two target puffs immediately one after other. The target puffs were timed with a 150 ms SOA from the previous, non-target air puffs. The non-target air puffs, which were 10 out of 11 of all the puffs, were the ones delivered between the index and middle fingers. The non-target puffs were timed on a random basis with five different SOAs, 433 ms, 678 ms, 899 ms, 1216 ms or 1601 ms from the previous air puff. This random timing of the air puffs made it impossible for the participants to know when the next air puff would occur and thus increased the attentive demands of noticing the occurrence of the target puffs. The air puffs were accompanied by a click sound but this sound was similar in target and non-target puffs. The two kinds of air puffs could only be distinguished by their location in the subject’s hand and possibly by the more immediate appearance of the target puffs after the previous non-target puff.

2.2.3. The task

The subjects were told their task was to count the target puffs of air delivered between their little finger and ring finger of their right hand and they were instructed to ignore the puffs of air delivered between their index finger and middle finger of the same hand. Before the experiment, the subjects had an opportunity to practice counting the air puffs during a short demonstration of 77 seconds in duration, during which puffs of air were presented without accompanying audiovisual stimuli. Before the start of the experimental trials with data collection, the subjects were instructed to keep their eyes on the screen where the black bars appeared but they were told they didn’t otherwise need to pay attention to the bars and tones. Excessive movement and eye blinks were discourager in order to avoid artifacts.

2.3. EEG recording

Electroencephalogram was recorded using Brainvision Recorder program and 28 Easy-Cap Ag/AgCl electrodes. Linked left and right mastoid were used as the reference electrode. Electrode impedances were at the level of 10 kΩ. Scalp potentials were amplified by Brainvision Quickamp amplifier, digitized with a sampling rate of 1000 Hz and filtered online with a band-pass of 0.1-100 Hz. Eye movements were monitored with a bipolar electrode above the left eye and another at the outer canthus of the same eye.

2.4. Data analysis

Offline EEG analysis was conducted by Brainvision Analyzer program. The electroencephalogram was segmented into 600 ms sweeps (including 110 ms prestimulus period) separately for deviant stimuli and for standard stimuli immediately preceding the deviants. The sweeps were filtered with a 0.1 Hz (slope of 24 dB/octave) high pass filter and a 30 Hz (slope of 24 dB/octave) low pass filter. Trials containing an artifact (a voltage exceeding ±100 μV at any electrode location) were excluded from the analysis. On average, 88 % of both the standard and the deviant trials were included in the analysis, the lowest percentages for an individual subject being 67 % of the standard trials and 68 % of the deviant ones. Next, the remaining sweeps were baseline corrected against their average during a 100 ms prestimulus period between 110 ms and 10 ms before the stimulus onset. Finally, the sweeps were averaged into ERP waveforms, and grand averaged waveforms across all subjects were created.

Based on visual scrutiny of the grand averaged ERP waveforms (Figure 1, Figure 2) three ERP peaks with noticeable differences between waveforms evoked by standard and deviant stimuli, were selected for further analysis. The first two ERP peaks, one around 100 ms and other around 170 ms after the stimulus onset, both occurred at a latencies possible for mismatch negativity (Näätänen, 2001). The third ERP peak, around 340 ms, occurred at a latency possible for P3a (Comerchero and Polich, 1999). 12 electrodes (Fz, F3, F4, Cz, C3, C4, Pz, P3, P4, Oz, O1, O2), in which the data quality for all the subjects was relatively good, were chosen for the statistical analysis. Means of the voltages over 20 ms time period around the ERP peaks were calculated.

Statistical analyses were performed by SPSS for Windows program. The mean voltages for the 20 ms time periods around the three separate ERP peaks were analyzed using separate repeated-measures analyses of variance (ANOVAs) with factors for stimulus type (standard vs. deviant), electrode anteriority (frontal vs. central vs. parietal vs. occipital) and electrode hemisphere (left vs. right vs. midline). As instructed by Nissinen (2003), Greenhouse-Geisser adjusted degrees of freedom and P values were used whenever Greenhouse-Geisser ε was 0.75 or lower; otherwise Huynh-Feldt adjusted degrees of freedom and P values were used. Paired samples t-tests were performed in order to further investigate the repeated measures ANOVA results whenever a significant main effect for the stimulus type or a significant interaction between the stimulus type and any of the other factors was found.

3. RESULTS

3.1. Behavioral task

Because of the randomness of the somatosensory stimuli, the correct answer for the number of the target air puffs was 78±1. After the experiment, seven out of ten participants gave an answer in this range. The remaining three participants gave answers 76, 73 and 88 (97 %, 94 % and 113 % of the correct count of 78).

3.2. ERP results

3.2.1. 90-110 ms poststimulus period

A repeated measures ANOVA revealed a significant main effect for the electrode anteriority, $F(1.568,14.112)=25.870$, $P<0.001$, indicating that the voltages in the

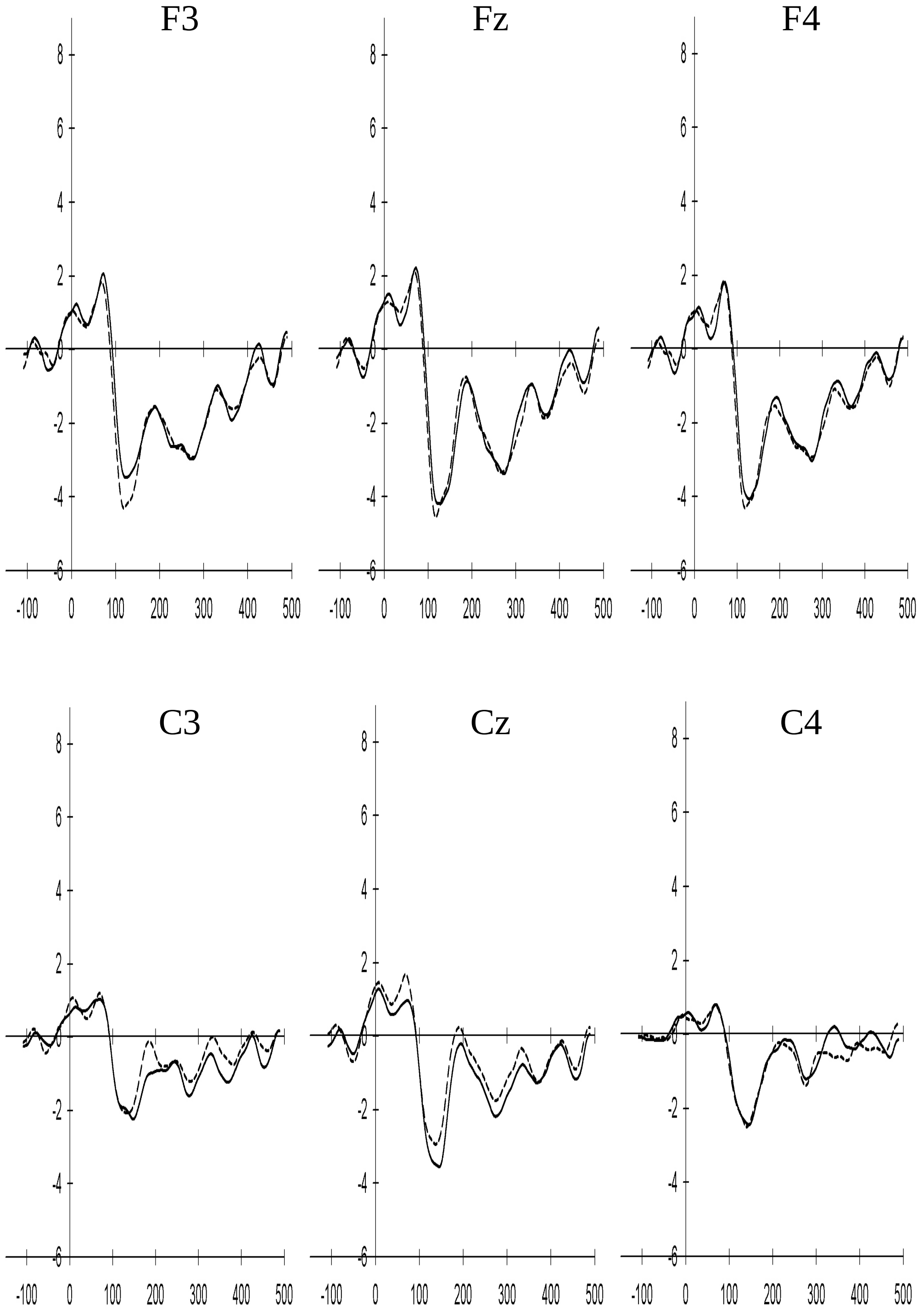

Figure 1: *The grand-averaged ERP waveforms at the frontal and the central electrodes. On the x-axis is the time relative to the stimulus onset in milliseconds. On the y-axis are the potentials in microvolts. The potentials for standard stimuli are indicated with dashed lines, the potentials for deviant stimuli, with solid lines.*

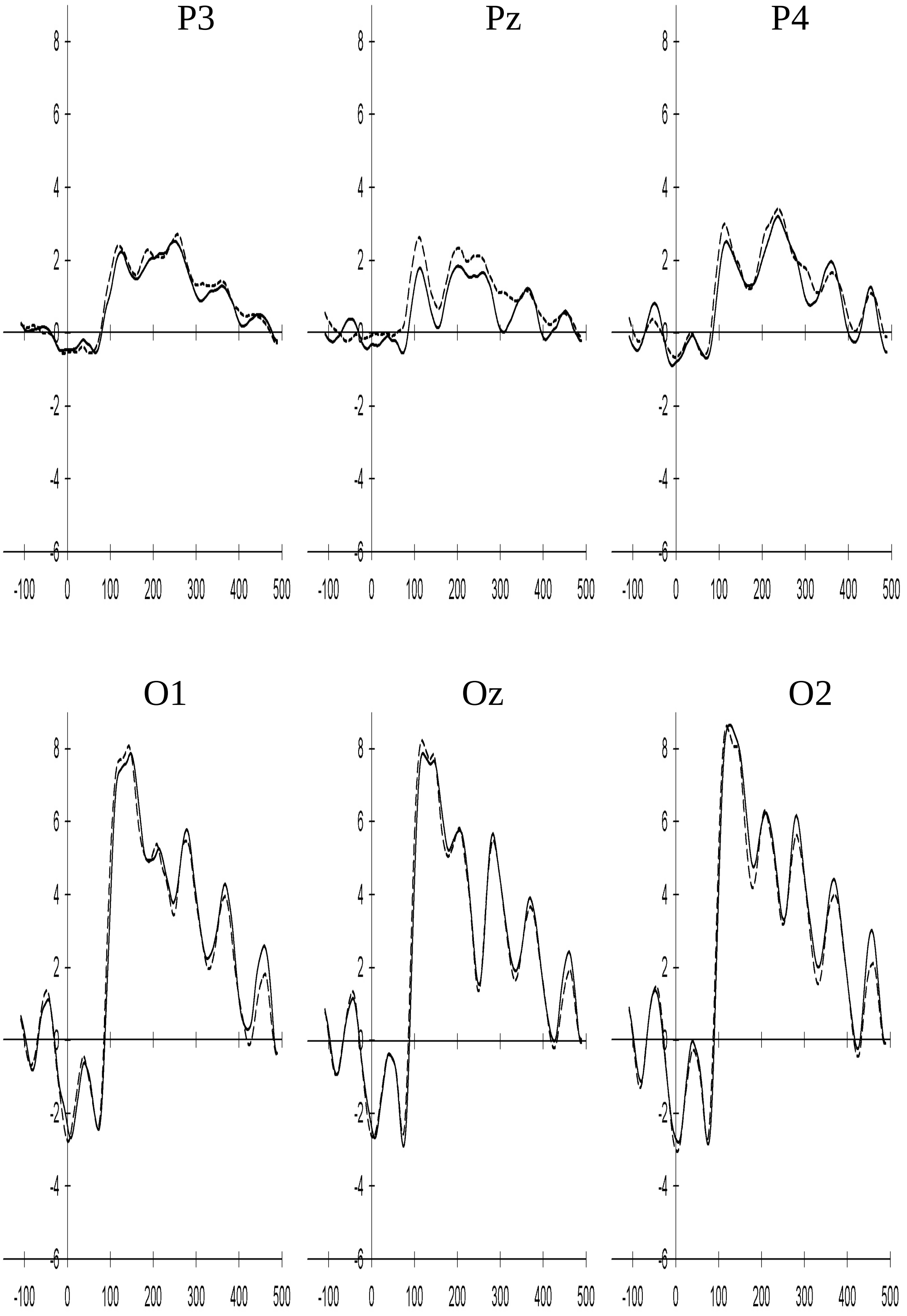

Figure 2: *The grand-averaged ERP waveforms at the parietal and the occipital electrodes. On the x-axis is the time relative to the stimulus onset in milliseconds. On the y-axis are the potentials in microvolts. The potentials for standard stimuli are indicated with dashed lines, the potentials for deviant stimuli, with solid lines.*

anterior electrodes were more negative than in the posterior electrodes. Also a significant interaction between the stimulus type and the electrode anteriority, F(1.160,10.440)=5.014, *P*=0.044, was found. None of the other main effects or interactions were significant.

Paired samples t-tests revealed a significant difference between the voltages to the standard and the deviant stimuli in the frontal electrodes, t(9)=-2.363, *P*=0.042, indicating more positive voltages to the deviant stimuli, and in the parietal electrodes, t(9)=2.516, *P*=0.033, indicating more negative voltages to the deviant stimuli. In the central and in the occipital electrodes, there were no significant differences between the voltages to the standard and the deviant stimuli.

3.2.2. 160-180 ms poststimulus period

A repeated measures ANOVA revealed a significant main effect for the electrode anteriority, F(1.032,9.286)=11.412, *P*=0.008, indicating that the voltages in the anterior electrodes were more negative than in the posterior electrodes. Also a significant interaction between the stimulus type and the electrode hemisphere, F(2.000,18.000)=6.491, *P*=0.008, was found. None of the other main effects or interactions were significant.

Paired samples t-tests revealed a significant difference between the voltages to the standard and the deviant stimuli in the left hemisphere electrodes, t(9)=3.475, *P*=0.007, indicating more negative voltages to the deviant stimuli, and in the midline electrodes, t(9)=2.813, *P*=0.020, also indicating more negative voltages to the deviant stimuli. In the right hemisphere electrodes, there was no significant difference between the voltages to the standard and the deviant stimuli.

3.2.3. 330-350 ms poststimulus period

A repeated measures ANOVA revealed a significant main effect for the electrode anteriority, F(1.188,10.690)=15.197, *P*=0.002, indicating that the voltages in the anterior electrodes were more negative than in the posterior electrodes. Also a significant interaction between the stimulus type and the electrode hemisphere, F(1.230,11.068)=4.904, *P*=0.043, and a significant three-way interaction between the stimulus type, the electrode hemisphere and the electrode anteriority, F(2.779,25.013)=5.213, *P*=0.007, were found. None of the other main effects or interactions were significant.

Paired samples t-tests revealed a significant difference between the voltages to the standard and the deviant stimuli in one electrode, C4, t (9)=-2.705, *P*=0.024, indicating more positive voltages to the deviant stimuli. There was no significant difference between the voltages to the standard and the deviant stimuli in any other electrode.

4. DISCUSSION

The aim of this study was to establish whether rare audiovisual feature conjunctions, differing from the standard conjunctions only in the probability of occurrence of the combinations themselves and not in the probability of occurrence of any constituent feature, are detected and analyzed by the brain's MMN generating system in the absence of attention. Indeed, the deviants elicited differential ERPs at three time periods: 90-110, 160-180, 330-350 ms. At the first latency, a positive difference frontally and a negative one parietally were found. At the second latency, the difference was negative at the left hemisphere and at the midline electrodes. At the third period, there was a positive difference at one electrode, C4.

This pattern of results indicates that differences between the standards and the deviants were really processed by the brains of the participants. However, this is also a more complex pattern of results than the single MMN expected, and it is necessary to make sense of these results. The first two of these peaks, at the 90-110 ms and the 160-180 ms poststimulus time windows, are negative polarity deflections at latencies the MMN has been reported at in the literature. The third differential peak at 330-350 ms period has a latency at which the MMN is rarely reported and is of positive polarity, which makes it a very improbable candidate for an MMN. This raises the question: Which one of the two first peaks is the MMN? It is possible that there is no one single MMN but there are two different processes that produce two separate MMN-like peaks. This kind of double mismatch processes have been reported both in auditory and visual experiments.

In the auditory domain, Zachau, Rinker, Körner, Kohls, Maas, Hennighausen, and Schecker (2005) discussed the late mismatch negativity (lMMN), which follows the earlier, more commonly known MMN and, according to the authors, reflects the process of transferring newly extracted rules about auditory stimuli into long term memory. However, their lMMN occurred at around 340 ms after the stimulus onset, which is a much later latency than the two negativities obtained in the current experiment.

In visual experiments, reports of double mismatch processes resembling the current results are not rare. For example, in the visual part of the feature combination study of Winkler et al. (2005) the deviant conjunctions elicited a negative difference wave shortly after the latency of 100 ms. This was followed by a positive difference wave peaking before the latency of 200 ms. Furthermore, Astikainen, Ruusuvirta, and Korhonen (2000) reported similar kind of results in an animal study in an experiment measuring event-related potentials to oddball stimuli varying in bar orientation. The measuring electrodes were placed in the visual cortex, the cerebellar cortex and the hippocampal dentate gyrus of rabbits. In the visual cortex, there was only one significant negative difference between ERPs to deviants and standards. However, in the cerebellar cortex there where two MMN-like positive difference waves at the latencies 75-150 ms and 175-200 ms and in the dentate gyrus, there where three similar difference waves at 25-50 ms, 75-100 ms and 150-175 ms. Also results with two cortical negativities have been reported. Kremlácek, Kuba, Kubová, and Langrová (2006) studied the effects of a deviant direction of motion of a visual stimulus in humans. They reported two negative difference waves, one peaking at 80 ms and the other prevailing in 145-265 ms poststimulus period.

In addition to the possibility of some kind of double MMN process, the other possibility is that of the two negativities at 90-110 ms and 160-180 ms periods, one is the MMN and the other is elicited by some other, yet unknown process. Considering the relative difficulty of the preattentive detection of the rare combinations among stimuli with equiprobable features on both varying stimulus dimensions, presented at a high stimulus-onset-asynchrony, it would be reasonable to assume that the later of these two, the 160-180 ms peak, is the more probable candidate for a single MMN. Unfortunately, because the phenomenon has received little research, no consensus exists about what kind of topography the audiovisual MMN should have. Thus, it can only be concluded that an MMN was elicited but it remains unclear whether the MMN was the first or the second negativity obtained or whether both negativities were part of some kind of double mismatch process.

As already mentioned in the Introduction, the methodology of the current study was designed to completely eliminate the possible influences of refractoriness on the MMN. This was achieved by presenting both the auditory frequencies and both the visual orientations at equal probability in both the standards and the deviants. Thus, the subjects' auditory pathways could not have passively adapted to any physical stimulus feature any more than to the other ones. This means that the differential ERPs obtained must be due to memory based stimulus processing.

The state of the participants' attention was controlled with a behavioral task in the somatosensory modality. The air puffs were delivered with varying intervals. This required constant attention from the subjects, as they had no possibility of anticipating when the next air puff was to be expected. This suggests that the subjects really were attending the somatosensory stimuli. However, there is an ERP component that can be used to assess whether the task-irrelevant audiovisual stimuli managed to capture the participants' attention: the P3a. This component is elicited by deviant stimuli that trigger an attention switch, and it has a frontocentral topography. In the current study, the third time window at 330-350 ms poststimulus period gave significantly more positive ERPs to deviant than to standard stimuli at one electrode, the C4. This is hardly the P3a component. The P3a should have a topographical distribution at many frontocentral electrodes (Spencer, Dien, & Donchin, 2001). Thus, it appears that the observed positivity was produced either by some unknown process or by pure chance. This implies that the audiovisual stimuli really were not attended.

The results of the current study, with an MMN elicited by unattended deviant audiovisual conjunctions, are in line with previous reports of audiovisual MMNs in the context of the McGurk (Colin et al. 2002b; Sams et al. 1991) and the ventriloquist illusions (Colin et al. 2002a; Stekelenburg et al. 2004).

Outside these illusions, there have been two studies reporting audiovisual MMNs in experimental settings in which the participants did not attend the stimuli of experimental interest. Besle et al. (2005) reported that audiovisual double deviants deviating from the standards with both the auditory and the visual features, produced an MMN that did not equal in amplitude the sum of the MMNs to the two kinds of deviants with the deviation in only one of these two stimulus dimensions. Rahne et al. (2007), in turn, demonstrated that unattended visual stimuli can affect the segregation of auditory stimuli into streams and determine, whether a deviant tone in stream elicits an MMN. The current study, using a methodology differing from these two studies, suggests that deviant audiovisual stimuli indeed elicit an MMN. In the current study, the participants performed a somatosensory task during the experiment, which is a more robust method of distracting the participants' attention than a task in one of the sensory modalities of experimental interest used by both Besle et al. and Rahne et al.

The current study further differed from the one of Besle et al. (2005) also in the method of preventing refractoriness. While Besle et al. had a standard frequency and a standard visual form, which were exchanged with the deviant features in half of blocks, in the current study both frequencies and both orientations were equiprobable all the time. There also appears to be a methodological difference in the data analysis. In the current study, only the standard stimuli immediately preceding deviants were taken into analysis, which resulted in the comparison of deviants and standards that were equally numerous. Besle et al., on the other hand, do not mention any correction of this kind, as they do not describe in great detail how they compared the MMN amplitude to the audiovisual double deviants with the sum of MMN amplitudes to the auditory and visual single deviants.

Despite the methodological differences, the current study and Besle et al. (2005) ended up having similar kind of results. In the current study, at the 160-180 ms poststimulus period, the amplitudes to deviant combinations were more negative than the amplitudes to standard ones at the left hemisphere and midline electrodes. In a similar fashion, Besle et al. reported that the additivity between the MMN amplitudes to double deviants and the MMN amplitudes to the visual and auditory single deviants, was violated at the 178-218 ms poststimulus period, at several left hemisphere parietotemporal electrodes. Together, these results seem to indicate that the left hemisphere has a special role in integrating unattended audiovisual stimulus combinations.

Also the study of Rahne et al. (2007) differed metholigically from the current study. The method for controlling the participants' attention was already mentioned. Furthermore, Rahne et al. used the frequency as the auditory dimension and the size of the visual stimulus as the visual dimension. The standards and the deviants were not distinguished by a rare combination of the visual and auditory features as was in the current study, but by a violation of regularity in the auditory sequence. This violation was detected by the MMN generating system when the visual stimuli were in accordance with the auditory stream stimuli following the regular sequence. In another experimental condition, the visual stimuli were in accordance with an irrelevant, different feature, the sound amplitude. In this experimental condition, occasional louder sounds interspersed in both streams where marked visually with no visual sign suggesting any kind of sound segregation into two different streams. In this condition, the two auditory streams indeed were not segregated by the participants and no MMN was elicited by the deviants in one stream. Therefore, the study of Rahne and the current one resembled each other in the fact that in neither study was there a single feature that was always either the standard or the deviant, but the regularities of the audiovisual presentation, although very different in the two studies, were the key of distinguishing the standards from the deviants in both studies. This suggests that the MMN generating system is capable of making sense of audiovisual regularities in a wide variety of experimental conditions.

The results reported by Rahne et al., however, do not have a great resemblance to the results of the current study in their topography. Rahne et al., namely, report significant differences at Fz, Cz, and Pz electrodes, with no differences at mastoids or at the parietal Oz electrode. Thus the MMN negativity of their study is not topographically very different from a normal auditory MMN.

The fact that the audiovisual feature conjunctions are processed in the absence of attention indicates that the model of a preattentive feature analysis and an attentive integration of these features, proposed by the FIT, does not apply to audiovisual processing. Even the version 2 (Treisman, 1988) of the FIT, which allows an initial conjunction detection at the level of the master map of locations before the analysis of the single features, does not explain the results of the current study and those of the earlier (Besle et al. 2005; Rahne et al. 2007) audiovisual MMN studies, as the analysis of probabilities of occurrence of the feature combinations really is beyond the notion of passive, unanalytic conjunction detection endorsed by the version 2 of the FIT. Thus, both the auditory (as discussed in the Introduction) and the audiovisual feature integration by the MMN generating system strongly challenge the generalizability of the FIT outside the visual modality. Even in the visual modality, which the FIT makes its claims about, there is one report of feature binding by the MMN generating system, demonstrating that rare conjunctions of orientation and color elicit an MMN no matter whether they were attended or not (Winkler et al. 2005). Consequently, the results of the current study strongly suggest that more research on preattentive visual feature binding needs to be carried out.

Besides the need to study visual feature binding, the current study also raises other possible lines of future research. The current study reported an MMN to deviant combinations of sound frequency and visual orientation and Besle et al. (2005) studied combinations of frequency and visual form. Rahne et al. (2007) reported an MMN to combinations of frequency and visual stimulus size that did not match the general sequence. This means that all these publications have dealt with combinations of frequency and a visual dimension. It seems possible that also features on other auditory dimensions besides frequency (timbre, intensity, duration, location, direction of movement) can be bound together with features on a variety of visual dimensions (color, brightness, orientation, form, size, duration, location, direction of movement) by the MMN generating system. However, this remains to be demonstrated by future research. Moreover, it still remains to be demonstrated that it is not only adult human subjects that can bind these audiovisual conjunctions together preattentively. This is why studies with newborn infants and ones with non-human animals are needed. Furthermore, there remains an exciting possibility of recording an MMN to deviant combinations of either visual or auditory stimulus components paired with somatosensory or olfactory components. This would introduce new sensory modalities that have been subject of few MMN studies, into research literature on bimodal binding by the MMN generating system.

REFERENCES

Astikainen, P., Ruusuvirta, T., Korhonen, T. (2000). Cortical and subcortical event-related potentials to oddball stimuli in rabbits. *Neuroreport*, 11, 1515-1517.

Astikainen, P., Ruusuvirta, T., Wikgren, J., & Korhonen, T. (2004). The human brain processes visual changes that are not cued by attended auditory stimulation. *Neuroscience Letters*, 368, 231-234.

Astikainen, P., Ruusuvirta, T., Wikgren, J., & Penttonen, M. (2006). Memory-based detection of rare sound feature combinations in anesthetized rats. *Neuroreport*, 17, 1561-1564.

Besle, J., Fort, A., & Giard, M. H. (2005). Is the auditory sensory memory sensitive to visual information? *Experimental Brain Research*, 166, 337-344.

Carrasco, M., & McElree, B. (2001). Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences of the United States of America*, 98, 5363-5367.

Carrasco, M., & Yeshurun, Y. (1998). The contribution of covert attention to the set-size and eccentricity effects in visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 673-692.

Clark, S. E., & Gronlund, S. D. (1996). Global matching models of recognition memory: How the models match the data. *Psychonomic Bulletin & Review*, 3, 37-60.

Colin, C., Radeau, M., Soquet, A., Dachy, B., & Deltenre, P. (2002a). Electrophysiology of spatial scene analysis: the mismatch negativity (MMN) is sensitive to the ventriloquism illusion. *Clinical Neurophysiology*, 113, 507-518.

Colin, C., Radeau, M., Soquet, A., & Deltenre, P. (2004). Generalization of the generation of an MMN by illusory McGurk percepts: voiceless consonants. *Clinical Neurophysiology*, 115, 1989-2000.

Colin, C., Radeau, M., Soquet, A., Demolin, D., Colin, F., & Deltenre, P. (2002b). Mismatch negativity evoked by the McGurk-MacDonald effect: a phonetic representation within short-term memory. *Clinical Neurophysiology*, 113, 495-506.

Comerchero, M. D., & Polich, J. (1999). P3a and P3b from typical auditory and visual stimuli. *Clinical Neurophysiology*, 110, 24-30.

Czigler, I., Balázs, L., & Winkler, I. (2002). Memory-based detection of task-irrelevant visual changes. *Psychophysiology*, 39, 869-873.

Czigler, I., Weisz, J., Winkler, I. (2006). ERPs and deviance detection: Visual mismatch negativity to repeated visual stimuli. *Neuroscience Letters*, 401, 178-182.

de Gelder, B., Böcker, 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.

Duncan, J., & Humphreys, G. W. (1989). Visual search and stimulus similarity. *Psychological Review*, 96, 433-458.

Dyson, B. J., & Quinlan, P. T. (2003). Feature and conjunction processing in the auditory modality. *Perception & Psychophysics*, 65, 254-272.

Dyson, B. J., & Quinlan, P. T. (2004). Stimulus processing constraints in audition. *Journal of Experimental Psychology: Human Perception and Performance*, 30, 1117-1131.

Gomes, H., Bernstein, R., Ritter, W., Vaughan, H. G. Jr., & Miller, J. (1997). Storage of feature conjunctions in transient auditory memory. *Psychophysiology*, 34, 712-716.

Hall, M. D., Pastore, R. E., Acker, B. E., & Huang, W. (2000). Evidence for auditory feature integration with spatially distributed items. *Perception & Psychophysics*, 62, 1243-1257.

Hari, R. (1994). Hermoston biosähköisten ja magneettisten ilmiöiden perusteet. In H. Lang, V. Häkkinen, A. Larsen, J. Partanen, & U. Tolonen (Ed.), *Sähköiset aivomme,* (pp. 17-22). Turku: Painokiila.

Houck, M. R., & Hoffman, J. E. (1986). Conjunction of color and form without attention: Evidence from an orientation-contingent color-aftereffect. *Journal of Experimental Psychology: Human Perception and Performance*, 12, 186-199.

Huang, L., & Pashler, H. (2005). Attentional capacity and task difficulty in visual search. *Cognition*, 94, B101-B111.

Humphreys, M. S., Pike, R., Bain, J. D., & Tehan, G. (1989). Global matching: A comparison of the SAM, Minerva II, Matrix and TODAM models. *Journal of Mathematical Psychology*, 33, 36-67.

Jamieson, R. K., Thompson, W. F., Cuddy, L. L., & Mewhort, D. J. (2003). Do conjunction errors in auditory recognition imply feature migration? *Canadian Journal of Experimental Psychology*, 57, 125-130.

Kremlácek, J., Kuba, M., Kubová, Z., & Langrová, J. (2006). Visual mismatch negativity elicited by magnocellular system activation. *Vision Research*, 46, 485-490.

McCollough, C. (1965). Color adaptation of edge-detectors in the human visual system. *Science*, 149, 1115-1116.

McElree, B., & Carrasco, M. (1999). The temporal dynamics of visual search: Evidence for parallel processing in feature and conjunction searches. *Journal of Experimental Psychology: Human Perception and Performance*, 25, 1517-1539.

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

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

Mondor, T. A., Zatorre, R. J., & Terrio, N. A. (1998). Constraints on the selection of auditory information. *Journal of Experimental Psychology: Human Perception and Performance*, 24, 66-79.

Mordkoff, J. T., & Yantis, S. (1991). An interactive race model of divided attention. *Journal of Experimental Psychology: Human Perception and Performance,* 17, 520-538.

Mordkoff, J. T., Yantis, S., & Egeth, H. E. (1990). Detecting conjunctions of color and form in parallel. *Percertion & Psychophysics*, 48, 157-168.

Möttönen, R., Krause, C. M, Tiippana, K., & Sams, M. (2002). Processing of changes in visual speech in the human auditory cortex. *Cognitive Brain Research*, 13, 417-425.

Nissinen, K. (2003). Toistomittausten analyysi. In H. Högmander, A. Kankainen, R. Lehtonen, E. Leskinen, A.L. Lyyra, K. Nissinen, & H. Oja, *Tilastolliset analyysimenetelmät, TILA04 Tilastomenetelmien jatkokurssi*, 3rd edition, (pp. 185-223). Jyväskylä.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology*, 38, 1-21.

Näätänen, R (2003). Mismatch negativity: clinical research and possible applications. *International Journal of Psychophysiology*, 48, 179-188.

Näätänen, R., Gaillard, A. W., & Mäntysalo, S. (1978). Early selective-attention effect on evoked potential reinterpreted. *Acta Psychologica*, 42, 313-329.

Näätänen, R., Jacobsen, T., & Winkler, I. (2005). Memory-based or afferent processes in mismatch negativity (MMN): a review of the evidence. *Psychophysiology*, 42, 25-32.

Näätänen, R., & Michie, P. T. (1979). Early selective-attention effects on the evoked potential: a critical review and reinterpretation. *Biological Psychology*, 8, 81-136.

Näätänen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., & Winkler, I. (2001). 'Primitive intelligence' in the auditory cortex. *Trends in Neurosciences*, 24, 283-288.

Paavilainen, P., Arajärvi, P., & Takegata, R. (2007). Preattentive detection of nonsalient contingencies between auditory features. *Neuroreport*, 18, 159-163.

Paavilainen, P., Simola, J., Jaramillo, M., Näätänen, R., & Winkler, I. (2001). Preattentive extraction of abstract feature conjunctions from auditory stimulation as reflected by the mismatch negativity (MMN). *Psychophysiology*, 38, 359-365.

Pazo-Alvarez, P., Amenedo, E., & Cadaveira, F. (2004). Automatic detection of motion direction changes in the human brain. *European Journal of Neuroscience*, 19, 1978-1986.

Pazo-Alvarez, P., Cadaveira, F., & Amenedo, E. (2003). MMN in the visual modality: a review. *Biological Psychology*, 63, 199-236.

Quinlan, P. T. (2003). Visual feature integration theory: past, present and future. *Psychological Bulletin*, 129, 643-673.

Rahne, T., Böckmann, M., von Specht, H., & Sussman, E. S. (2007). Visual cues can modulate integration and segregation of objects in auditory scene analysis. *Brain Research*, 1144, 127-135.

Ruusuvirta, T., & Huotilainen, M. (2004). The human brain processes repeated auditory feature conjunctions of low sequential probability. *Neuroscience Letters*, 355, 97-100.

Ruusuvirta, T., Huotilainen, M., Fellman, V., & Näätänen, R. (2003). The newborn human brain binds sound features together. *Neuroreport*, 14, 2117-2119.

Ruusuvirta, T., Huotilainen, M., Fellman, V., & Näätänen, R. (2004). Newborn human brain identifies repeated auditory feature conjunctions of low sequential probability. *European Journal of Neuroscience*, 20, 2819-2821.

Ruusuvirta, T., Koivisto, K., Wikgren, J., & Astikainen, P. (2007). Processing of melodic contours in urethane-anaesthetized rats. *European Journal of Neuroscience*, 26, 701-703.

Saarinen, J., Paavilainen, P., Schöger, E., Tervaniemi, M., & Näätänen, R. (1992). Representation of the abstract attributes of auditory stimuli in the human brain. *Neuroreport*, 3, 1149-1151.

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., Hämäläinen, M., Hari, R., Lounasmaa, O. V., Lu, S. T., & Simola, J. (1991). Seeing speech: visual information from lip movements modifies activity in the human auditory cortex. *Neuroscience Letters*, 127, 141-145.

Shalgi, S., & Deouell, L. Y. (2007). Direct evidence for differential roles of temporal and frontal components of auditory change detection. *Neuropsychologia*, 45, 1878-1888.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology*, 38, 343-358.

Stekelenburg, J. J., Vroomen, J., & de Gelder, B. (2004). Illusory sound shifts induced by the ventriloquist illusion evoke the mismatch negativity. *Neuroscience Letters*, 357, 163-166.

Surakka, V., Tenhunen-Eskelinen, M., Hietanen, J. K., & Sams, M. (1998). Modulation of human auditory information processing by emotional visual stimuli. *Cognitive Bran Research*, 7, 159-163.

Sussman, E., Gomes, H., Nousak, J. M., Ritter, W., & Vaughan, H. G. Jr. (1998). Feature conjunctions and auditory sensory memory. *Brain Research*, 793, 95-102.

Takegata, R., Brattico, E., Tervaniemi, M., Varyagina, O., Näätänen, R., & Winkler, I. (2005). Preattentive representation of feature conjunctions for concurrent spatially distributed auditory objects. *Cognitive Brain Research*, 25, 169-179.

Takegata, R, Huotilainen, M., Rinne, T., Näätänen, R., & Winkler, I. (2001). Changes in acoustic features and their conjunctions are processed by separate neuronal populations. *Neuroreport*, 12, 525-529.

Takegata, R., Paavilainen, P., Näätänen, R., & Winkler, I. (1999). Independent processing of changes in auditory single features and feature conjunctions in humans as indexed by the mismatch negativity. *Neuroscience Letters*, 266, 109-112.

Thompson, W. F., Hall, M. D., & Pressing, J. (2001). Illusory conjunctions of pitch and duration in unfamiliar tone sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 128-140.

Treisman, A. (1986). Features and objects in visual processing. *Scientific American*, 255, 106-115.

Treisman, A. (1988). Features and objects: the fourteenth Bartlett memorial lecture. *The Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 40A, 201-237.

Treisman, A. M., & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, 12, 97-136.

Treisman, A., & Schmidt, H. (1982). Illusory conjunctions in the perception of objects. *Cognitive Psychology*, 14, 107-141.

Ullsperger, P., Erdmann, U., Freude, G., & Dehoff, W. (2006). When sound and picture do not fit: Mismatch negativity and sensory interaction. *International Journal of Psychophysiology*, 59, 3-7.

Widdman, A., Kujala, T., Tervaniemi, M., Kujala, A., & Schröger, E. (2004). From symbols to sounds: Visual symbolic information activates sound representations. *Psychophysiology*, 41, 709-715.

Winkler, I., Czigler, I., Sussman, E., Horváth, J., & Balázs, L. (2005). Preattentive binding of auditory and visual stimulus features. *Journal of Cognitive Neuroscience*, 17, 320-339.

Woods, D. L., Alain, C., Diaz, R., Rhodes, D., & Ogawa, K. H. (2001). Location and frequency cues in auditory selective attention. *Journal of Experimental Psychology: Human Perception and Performance*, 27, 65-74.

Woods, D. L., Alain, C., & Ogawa, K. H. (1998). Conjoining auditory and visual features during high-rate serial presentation: Processing and conjoining two features can be faster than processing one. *Perception & Psychophysics*, 60, 239-249.

Zachau, S., Rinker, T., Körner, B., Kohls, G., Maas, V., Hennighausen, K., & Schecker, M. (2005). Extracting rules: early and late mismatch negativity to tone patterns. *Neuroreport*, 16, 2015-2019.