available at www.sciencedirect.com
ScienceDirect
www.elsevier.com/locate/brainres
ELSEVIER
BRAIN RESEARCH

Research Report

Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension

Roel M. Willems[a,*], *Robert Oostenveld*[a], *Peter Hagoort*[a,b,c]

[a]*F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, P.O. Box 9101, 6500 HB Nijmegen, The Netherlands*
[b]*Max Planck Institute for Psycholinguistics, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands*
[c]*Nijmegen Institute for Cognition and Information, Radboud University Nijmegen, P.O. Box 9104, 6500 HE Nijmegen, The Netherlands*

ARTICLE INFO

Article history:
Accepted 16 April 2008

Keywords:
Language
Visual
EEG
Oscillation
Frequency
Meaning

ABSTRACT

Language is often perceived together with visual information. This raises the question on how the brain integrates information conveyed in visual and/or linguistic format during spoken language comprehension. In this study we investigated the dynamics of semantic integration of visual and linguistic information by means of time-frequency analysis of the EEG signal. A modified version of the N400 paradigm with either a word or a picture of an object being semantically incongruous with respect to the preceding sentence context was employed. Event-Related Potential (ERP) analysis showed qualitatively similar N400 effects for integration of either word or picture. Time-frequency analysis revealed early specific decreases in alpha and gamma band power for linguistic and visual information respectively. We argue that these reflect a rapid context-based analysis of acoustic (word) or visual (picture) form information. We conclude that although full semantic integration of linguistic and visual information occurs through a common mechanism, early differences in oscillations in specific frequency bands reflect the format of the incoming information and, importantly, an early context-based detection of its congruity with respect to the preceding language context.

1. Introduction

Language is often perceived in the presence of concomitant semantic information. For instance, linguistic utterances often take place with reference to objects in the environment. Consider someone showing his friend the features of a new car. The speaker will perhaps talk about improvements to the engine of the vehicle, while at the same time showing the engine to the listener. Here, the co-occurrence of language and visual information is an important feature of the way the message is conveyed by the speaker as well as how it is understood by the listener. A consequence of this common co-occurrence is that the brain continuously has to integrate streams of information conveyed through different modalities during language comprehension. Importantly, as in the example above, such integration has to happen at a semantic level. That is, there is no way in which the form properties of the visual information and of the spoken language overlap. Here we investigated the possibility that such integration can be distinguished neurally in terms of differences in changes in power in specific frequency bands.

Previous research indicates that integration of visual and linguistic information at this level of processing taxes over-

* *Corresponding author*. Fax: +31 243610989.
E-mail address: roel.willems@fcdonders.ru.nl (R.M. Willems).

0006-8993/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.brainres.2008.04.065

lapping neural correlates. Several studies employing the event-related potential (ERP) technique show that an incongruous picture of an object evokes a qualitatively similar N400 effect as compared to an anomalous word (the N400 is thought to reflect semantic integration load of an item with respect to a previous context; see below). For instance, Ganis et al. (1996) presented sentences that either ended with a word or a picture that could be anomalous or not. Similar N400 effects were found to anomalous words and pictures. The scalp distribution for the anomalous pictures was more frontal than for the anomalous words. Nigam and colleagues (Nigam et al., 1992) also found similar N400 effects for pictures and words, but did not find a difference in scalp distribution. However, this might be due to the limited number of electrodes that they recorded from, which did not cover the frontal part of the brain. Federmeier and Kutas (2001) found a correlation between the amplitude of the N400 effect and the semantic fit of a picture with respect to the preceding part of a sentence. Again, there was a frontal scalp distribution for the effects. Additionally, they observed an N300 effect to the anomalous pictures. Some other ERP studies have investigated the processing of visual information following a *visual* context instead of a language context. West and Holcomb (2002) for instance presented a series of pictures forming a simple story. The last picture was either a congruous or an incongruous ending of the story. Incongruous pictures elicited increased N300 and N400 effects, with a maximal distribution over centro-frontal electrodes. Sitnikova et al. (2003) had congruous or incongruous objects appear in video clips of real-world events. They observed an N400 effect for the incongruous objects with a fronto-central maximum in the scalp distribution. Ganis and Kutas (2003) had congruous or incongruous objects appear in still images of real-world events. An increased negativity strongly resembling the N400 was observed for the incongruous as compared to the congruous objects. Finally, in an earlier report of the ERP analysis of part of the data presented in this paper, we found that incongruous pictures and words evoke similar N400 effects and lead to overlapping activations in left inferior frontal cortex (Willems et al., in press; see also Özyürek et al., 2007; Willems et al., 2007).

In summary, ERP studies manipulating the semantic fit of pictures in relation to a (sentence) context report similar N400 amplitudes and onset latencies as found for integration of semantic information conveyed through a word. Moreover, integration of information from pictures and words into a sentence context leads to overlapping activations in left inferior frontal cortex.

From these and other findings it has been claimed that despite differences in representational format, integration of linguistic and non-linguistic semantic information with language recruits the same neural mechanisms. In line with this, Hagoort and colleagues showed that integration of two types of knowledge (lexical semantic knowledge and general knowledge of the world) during sentence comprehension follows the same neural time course and recruits an overlapping neural locus (Hagoort et al., 2004; Hagoort and van Berkum, 2007). However, it was also found that integration of these information types can be distinguished in terms of differences in frequency band power of the EEG (Hagoort et al., 2004). Specifically, the world knowledge violations led to an increase in gamma band power that was not observed for semantic violations. Here we investigated whether analogous to Hagoort et al., visual and linguistic information also elicit different responses in the frequency domain.

Time-frequency analysis can reveal effects that go unnoticed in the time-locked ERP, due to the averaging of the signal in ERP analysis. In several domains of cognitive neuroscience it has proven to be fruitful to study frequency-specific changes in power to specific cognitive events (see e.g. Engel et al., 2001; Tallon-Baudry, 2003; Herrmann et al., 2004b; Jensen et al., 2007). However, analysis in the time-frequency domain remains less well studied in the neurocognition of language (but see Bastiaansen and Hagoort, 2006 for a recent review).

To investigate the issue of frequency-specific effects related to linguistic and visual semantic processing during sentence comprehension, we employed the N400 paradigm. A word with a meaning that is incongruous with respect to a preceding part of the sentence leads to a more negative deflection in the ERP around 400 ms after word onset (Kutas and Hillyard, 1980). This effect is labeled the N400 effect and has become a well-established indicator of semantic integration of for instance a word into a preceding context (see Kutas and Van Petten, 1994 for review; Brown et al., 2000). In contrast, the oscillatory correlates of semantic processing are not well established. Semantic processing has been linked to increases in power in the theta band (around 4–6 Hz) by some (Bastiaansen et al., 2005) and by decreases in power in the alpha band (around 10 Hz) by others (Rohm et al., 2001).

Relevant for the present paper are two recent studies in which an N400 paradigm was used to assess oscillatory correlates of semantic processing during sentence comprehension (Hald et al., 2006; Davidson and Indefrey, 2007). Hald and colleagues observed an increase in theta band (3–5 Hz) power after a semantic incongruity. This was interpreted as reflecting an increased difficulty in lexical selection in the case of a semantically incongruous word (Hald et al., 2006). Interestingly, also an early (50–200 ms after presentation of the critical word) decrease in power in the gamma band (35–45 Hz) was observed. This effect was tentatively linked to the absence of integration or 'unification' at the sentence level in the incongruous condition. That is, in the case of a semantic incongruity unification of all words of the sentence into a coherent whole is rendered impossible, leading to the gradual built-up of gamma power to be halted (Hald et al., 2006).

Davidson and Indefrey (2007) observed a similarly late increase in theta power (3–7 Hz) after a semantic violation. No other differences were observed, but it should be noted that the gamma band was not analyzed in that study.

Here we investigated similarities or differences between oscillatory correlates of integration of information conveyed linguistically (words) or visually (pictures) during spoken sentence comprehension. To do this, we adapted the N400 paradigm to modulate semantic load of either a spoken word, a picture or of both. Participants listened to spoken sentences in which a critical word was presented which could be either semantically congruous or incongruous with respect to the preceding part of the sentence. Together with the critical word, a picture was presented which could also be congruous or incongruous (Table 1). There were four conditions: 1) Correct condition (Picture congruous, Word congruous) 2) Language

Table 1 – An example of the stimulus materials

Dutch: Voor in de keuken kocht zij een eenvoudige kom/**trein** en borden

English: For (use in) the kitchen she bought a simple bowl/**train** and plates

Correct condition

P+L+: For in the kitchen she bought a simple bowl and plates

Language mismatch

P+L-: For in the kitchen she bought a simple **train** and plates

Picture mismatch

P-L+: For in the kitchen she bought a simple bowl and plates

Double mismatch

P-L-: For in the kitchen she bought a simple **train** and plates

Pictures were displayed time-locked to the onset of the noun (underlined). Note that the condition coding (P + L+, P+L-, etc.) refers to the match/mismatch of either the noun (Language: L) or the Picture (Picture: P) to the part of the sentence preceding the word that is underlined, with a minus sign indicating a mismatch. That is, in the correct condition (P+L+), both the word 'bowl' as well as the picture [BOWL] fit the preceding sentence context. In the Language mismatch condition (P+L-), the word 'train' does not fit the preceding sentence context, whereas the picture [BOWL] does fit. Conversely, in the Picture mismatch condition (P-L+) the picture [TRAIN] does not fit the preceding sentence context, whereas the word 'bowl' does fit. Finally, in the Double mismatch condition (P-L-) both the word 'train' and the picture [TRAIN] do not fit the preceding sentence context. Mismatching words are indicated in bold. All stimuli were in Dutch; the literal translation in English is ungrammatical.

mismatch (Picture congruous, Word incongruous) 3) Picture mismatch (Picture incongruous, Word congruous) and 4) Double mismatch (Picture incongruous, Word incongruous). The Double mismatch condition was added to test whether effects are a reflection of increased semantic load with regard to the preceding sentence context or of mismatching co-occurring picture and word. That is, in the Language mismatch condition and the Picture mismatch condition one can argue that possible effects are driven by the fact that picture and word in these conditions convey a different meaning. If so, the effects would not be a reflection of sentence-level semantic integration. Since in the Double mismatch condition picture and word convey the same meaning (but are incongruous with respect to the preceding sentence context), this cannot be the case in this condition.

We hypothesized that all mismatch conditions would evoke an N400 effect in the ERP analysis, corroborating earlier findings as described above and as we have reported before (Willems et al., in press). Since the results of the ERP analysis of almost the same data set have been published and discussed elsewhere (Willems et al., in press), our focus will be on the outcome of analysis in the frequency domain. We hypothesized a relatively late theta band power increase to be a reflection of a general (that is, not language-specific) integration mechanism, analogous to the N400. If this is indeed the case, it should be obtained in all three mismatch conditions. Furthermore we expected decreases in the alpha (Rohm et al., 2001) and/or gamma (Hald et al., 2006) frequency bands in response to the Language mismatch condition. A crucial question was whether similar effects would be observed when the picture, but not the word was in discordance with the previous sentence context. Alternatively, effects specifically related to increased semantic load as conveyed through a visual stimulus may be observed. One candidate frequency band to manifest such specific effect is the gamma frequency band, which has been implicated in successful recognition of objects (see e.g. Rodriguez et al., 1999; Tallon-Baudry, 2003).

2. Results

2.1. *ERP results*

Although the focus of this paper is on the results obtained from the time-frequency analysis, we briefly report results of the ERP analysis to be able to compare the two measures. The ERP grand average waveforms (Fig. 1A) show a N1 followed by a P2, followed by a negativity resembling the N400. The latter negativity shows clear differences between the correct condition (black in Fig. 1A) and the mismatch conditions (colored in Fig. 1A), lasting from 300 ms to 750 ms after stimulus onset. Statistical analysis involved repeated measures analysis of variance (ANOVA) of the mean amplitude in the 300–600 ms time window with factors Condition (Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 scalp electrodes) (Table 2). There was a main effect of Condition ($F(3,45)=6.17$, MSe=61.75, $p=0.002$), but no Condition×Electrode interaction ($F(78, 1170)=1.09$, MSe=8.32, $p=0.364$). Planned comparisons showed that all mismatch conditions differed significantly from the Correct condition (Language mismatch vs. Correct condition: $F(1,15)=12.98$, MSe=154.66, $p=0.003$; Picture mismatch vs. Correct condition: $F(1,15)=5.61$, MSe=183.79, $p=0.032$; Double mismatch vs. Correct condition: $F(1,15)=5.21$, MSe=114.92, $p=0.037$). The Picture mismatch and Language mismatch conditions were not significantly different from each other ($F(1,15)=1.91$, MSe=84.52, $p=0.187$).

In all mismatch conditions the N400 effect was widely spread across the scalp (Fig. 1B), but with a more anterior distribution than is usually observed for the N400 effect to spoken or written words (e.g. Kutas and Hillyard, 1980; Hagoort and Brown, 2000; van den Brink et al., 2001). Such a relatively anterior distribution has been observed before in studies employing the N400 paradigm with pictures (e.g. Ganis et al., 1996; Federmeier and Kutas, 2001).

2.2. *Time-frequency results*

2.2.1. *Analysis with a priori defined time-frequency windows*

Statistical analysis involved repeated measures analyses of variance (ANOVA) on mean power in six pre-defined frequency-time windows of interest with factors Condition (Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 scalp electrodes). The frequency bands tested were based upon previous literature.

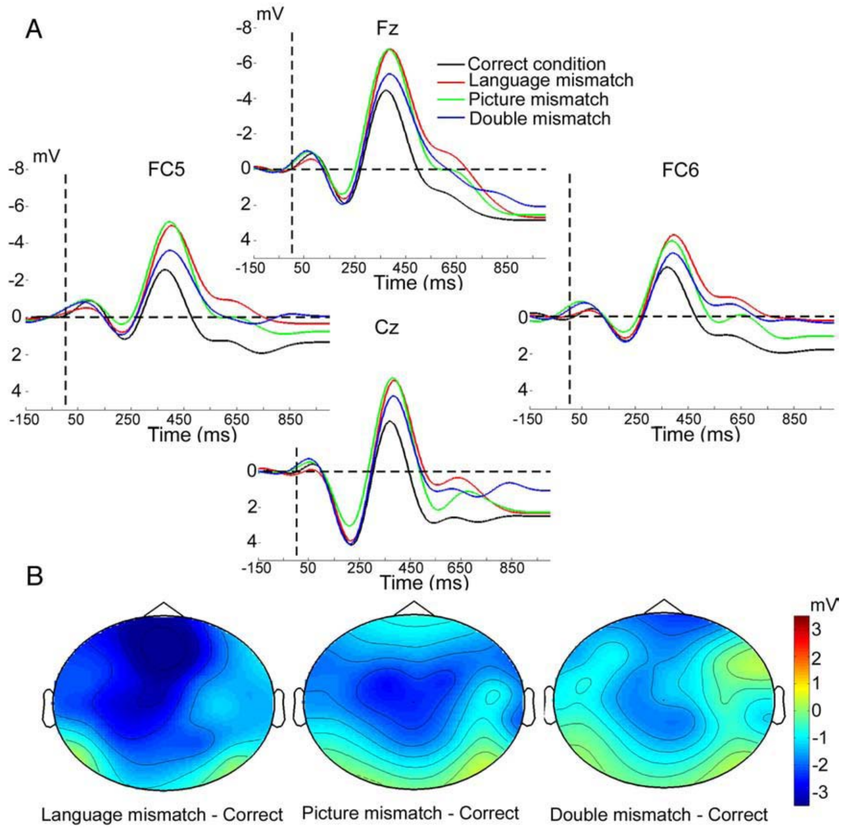

Fig. 1 – A) Averaged event-related potentials time-locked to the onset of the critical word. Presented are the waveforms from electrodes FC5 (left), Fz (upper), FC6 (right) and Cz (lower) of all four conditions. The increased negativity of the colored lines (Mismatch conditions) as compared to the black line (Correct condition) is clearly visible. Negative is plotted upwards. Waveforms are low-pass filtered for illustration purposes only. B) Scalp topographies of the N400 effects in the 300-600 ms range for the Language mismatch-Correct condition (left), Picture mismatch-Correct condition (middle) and Double mismatch-Correct condition (right) comparisons. Note the more anterior distribution than is normally observed for the N400 effect elicited by spoken or written words.

We tested for effects in the theta band (4–6 Hz) (Bastiaansen et al., 2005; Hald et al., 2006; Davidson and Indefrey, 2007), alpha band (8–12 Hz) (Rohm et al., 2001) and the lower gamma band (40–50 Hz) (Hagoort et al., 2004; Hald et al., 2006). Separate ANOVAs were conducted in early (0–300 ms) and late (350–750 ms) time windows. The results are summarized in Table 3. In the case of a main effect of Condition or Condition × Electrode interaction, planned comparisons were performed to test for differences between each mismatch condition vs the Correct condition as well as between the Picture mismatch condition and the Language mismatch condition. The factor Electrode was always significant and is not reported in the text (see Table 3). Fig. 2 shows Time-Frequency representations of each single condition; Fig. 3 shows Time-Frequency representations of differences between conditions.

In the early theta cluster (4–6 Hz, 0–300 ms) there was no significant main effect of Condition ($F(3,45)=1.13$, MSe = 82.39, $p=0.322$). The Condition × Electrode interaction was marginally significant ($F(78,1170)=1.839$, MSe = 10.89, $p=0.080$). Planned comparisons showed a trend towards statistical significance for a power increase in the Double mismatch vs. Correct condition comparison only ($F(1,15)=3.74$, MSe = 34.47, $p=0.072$).

In the late theta cluster (4–6 Hz, 0–300 ms) there was a main effect of Condition ($F(3,45)=4.70$, MSe = 20.10, $p=0.017$) but no Condition × Electrode interaction ($F(78,1170)=1.47$, MSe = 104.43, $p=0.183$). Planned comparisons revealed that all mismatch conditions evoked significantly stronger power as compared to the correct condition (Language mismatch vs. Correct condition: $F(1,15)=9.23$, MSe = 29.99, $p=0.008$; Picture mismatch vs. Correct condition: $F(1,15)=22.76$, MSe = 11.93, $p<0.001$; Double mismatch vs. Correct condition: $F(1,15)=9.44$, MSe = 20.46, $p=0.008$). Effects were maximal over Frontal electrode sites for all comparisons (Figs. 3A, B and C). The Language mismatch and Picture mismatch condition did not differ significantly from each other ($F<1$).

In the early alpha cluster (8–12 Hz, 0–300 ms), there was a marginally significant main effect of Condition ($F(3,45)=2.90$, MSe = 32.95, $p=0.071$) but no Condition × Electrode interaction ($F<1$). Planned comparisons showed that the Double mismatch condition elicited significantly less alpha power as compared to the Correct condition ($F(1,15)=4.54$, MSe = 13.19, $p=0.050$). The effect was maximal over centro-posterior electrodes (Fig. 3C). The Language mismatch condition was not statistically different from the Correct condition ($F(1,15)=$

Table 2 – Results of the ERP analysis

Factor	F	MSe	P
Condition	F(3,45) = 6.17	61.75	**0.002**
Electrode	F(26,390) = 15.98	353.02	**< 0.001**
Condition × Electrode	F(78,1170) = 1.09	8.32	0.364
Planned comparisons			
Language mismatch vs. Correct condition	F(1,15) = 12.98	154.66	**0.003**
Picture mismatch vs. Correct condition	F(1,15) = 5.61	183.79	**0.032**
Double mismatch vs. Correct condition	F(1,15) = 5.21	114.92	**0.037**
Picture mismatch vs. Language mismatch	F(1,15) = 1.91	84.52	0.187

Analysis involved repeated measures analysis of variance (ANOVA) on the mean amplitude in the 300-600 ms time- window with factors Condition (Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 levels, see Experimental Methods section). Planned comparisons involved testing for differences between each mismatch condition and the Correct condition as well as between the Picture mismatch condition and the Language mismatch condition. Huynh-Feldt correction for violation of the sphericity assumption was applied (Huynh and Feldt, 1976), but original degrees of freedom are reported.

3.28, MSe=32.95, p=0.093), neither did the Picture mismatch condition differ from the Correct condition (F(1,15)=1.05, MSe=18.80, p=0.321). Importantly, however, the Language mismatch evoked significantly less alpha power as the Picture mismatch condition (F(1,15)=5.08, MSe=63.90, p=0.040). Again, the effect was maximal over centro-posterior electrodes (Fig. 3D).

In the late alpha cluster (8–12 Hz, 350–750 ms) there was no main effect of Condition (F(3, 45)=1.73, MSe=34.60, p=0.202) or a Condition×Electrode interaction (F<1).

In the early gamma cluster (40–50 Hz, 0–300 ms), there was a main effect of Condition (F(3,45)=3.63, MSe=0.28, p=0.023) but no Condition×Electrode interaction (F<1). Planned comparisons showed that only the Picture mismatch condition led to a significant decrease in power compared to the Correct condition (F(1,15)=4.87, MSe=0.36, p=0.043). Moreover, the Picture mismatch condition was significantly different from the Language mismatch condition (F(1,15)=13.62, MSe=0.41, p=0.002). The scalp distribution shows two maxima, one over left and one over right frontal electrodes, which is a rather unusual distribution (Fig. 3B). Inspection of the mean power differences of this comparison from the two electrodes in which the effect was maximal showed that the effect is consistent over participants in electrode FC6, and can therefore not be attributed to an artifact present in only some of the

Table 3 – Results of time-frequency analysis in a priori defined time-frequency clusters

Frequency	Factor	Early (0–300 ms)				Late (350–750 ms)		
		F	MSe	df	p	F	MSe	P
Theta (4–6 Hz)	Condition	1.13	82.39	3, 45	0.322	4.70	20.10	**0.017**
	Electrode	33.71	2902.45		**<0.001**	60.42	97.68	**<0.001**
	Cond × Electrode	1.839	10.89	78, 1170	0.080	1.47	104.43	0.183
	Planned comparisons							
	Language mismatch vs. Correct	2.21	87.86	1, 15	0.158	9.23	29.99	**0.008**
	Picture mismatch vs. Correct condition	<1		1, 15	ns	22.76	11.93	**<0.001**
	Double mismatch vs. Correct condition	3.74	34.47	1, 15	0.072	9.44	20.46	**0.008**
	Picture mismatch vs. Language mismatch	<1		1, 15	ns	<1		ns
Alpha (8–12 Hz)	Condition	2.90	32.95	3, 45	0.071	1.73	34.60	0.202
	Electrode	13.78	139.92	26, 390	**<0.001**	15.49	85.82	**<0.001**
	Cond × Electrode	<1		78, 1170	ns	<1		ns
	Planned comparisons							
	Language mismatch vs. Correct	3.28	32.95	1, 15	0.093			
	Picture mismatch vs. Correct condition	1.05	18.80	1, 15	0.32			
	Double mismatch vs. Correct condition	4.54	13.19	1, 15	0.050			
	Picture mismatch vs. Language mismatch	5.08	63.90	1, 15	**0.040**			
Gamma (40–50 Hz)	Condition	3.63	0.28	3, 45	**0.023**	1.07	0.27	0.365
	Electrode	3.01	45.02	26, 390	**0.013**	3.15	40.59	**0.008**
	Cond × Electrode	<1		78, 1170	ns	1.32	0.22	0.200
	Planned comparisons							
	Language mismatch vs. Correct	1.41	0.76	1, 15	0.254			
	Picture mismatch vs. Correct condition	4.87	0.36	1, 15	**0.043**			
	Double mismatch vs. Correct condition	<1		1, 15	ns			
	Picture mismatch vs. Language mismatch	13.62	0.41	1, 15	**0.002**			

Analysis involved repeated measures analysis of variance (ANOVA) on mean power in six pre-defined frequency-time windows of interest, chosen on the basis of previous literature (i.e. Rohm et al., 2001; Hagoort et al., 2004; Bastiaansen et al., 2005; Hald et al., 2006; Davidson and Indefrey, 2007). Data from the theta band (4-6 Hz), alpha band (8-12 Hz) and the lower gamma band (40-50 Hz) were analyzed in early (0-300 ms) and late (350-750 ms) time windows in ANOVAs with factors Condition (Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 levels, see Experimental procedures section). Planned comparisons involved testing for differences between each mismatch condition and the Correct condition as well as a direct comparison between Picture mismatch and Language mismatch conditions. Huynh–Feldt correction for violation of the sphericity assumption was applied (Huynh and Feldt, 1976), but original degrees of freedom are reported.

Please cite this article as: Willems, R.M., et al., Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension, Brain Res. (2008), doi:10.1016/j.brainres.2008.04.065

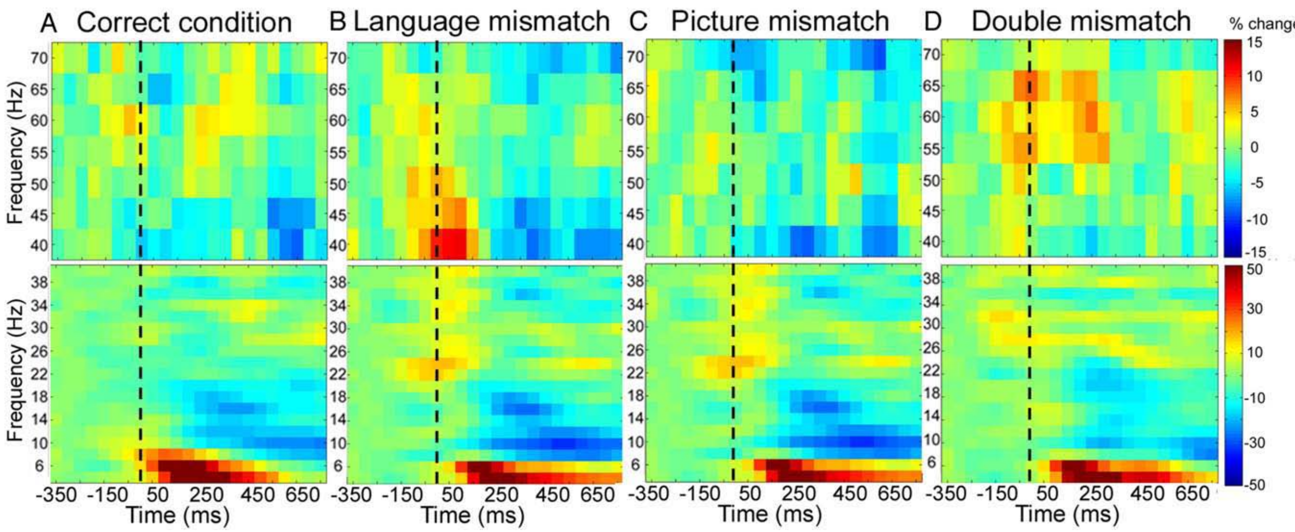

Fig. 2 – Time-frequency representations of all conditions. This figure shows power increases/decreases in response to critical word/picture (at time = 0) with respect to baseline in all analyzed frequency bands. Power is normalized with respect to power in the −350 to 0 ms time window in each frequency band by computing the relative change (percent signal change) as compared to the baseline condition for each frequency band separately. That is, baseline correction involved subtracting the mean of the baseline of that specific frequency band from the measured value and dividing this number by the mean power in the baseline (value − baseline/baseline). Therefore, the values in the figure represent percentage power change as compared to baseline. It was made sure that no post-stimulus activation was included in the baseline period due to conversion into the frequency domain. Although instructive, this figure does not clearly illustrate the differences between conditions. These are displayed in Fig. 3.

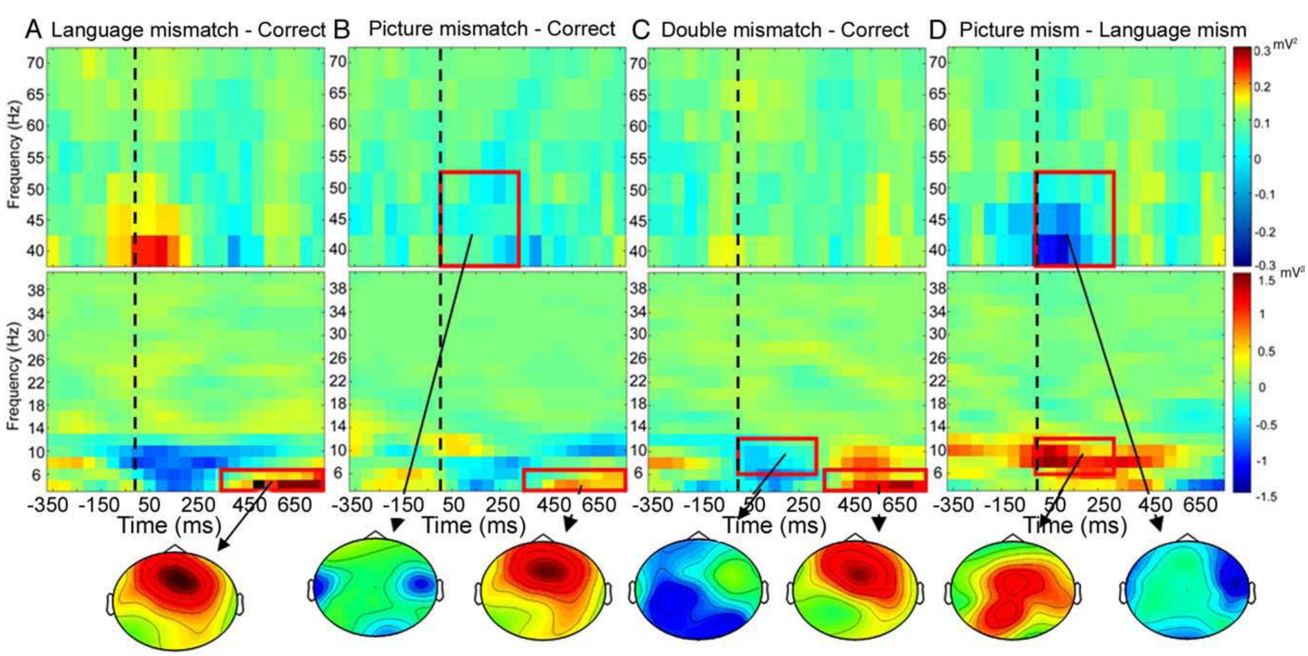

Fig. 3 – Average time-frequency representations of Language mismatch-Correct condition (A), Picture mismatch-Correct condition (B), Double mismatch-Correct condition (C) and Picture mismatch-Language mismatch condition (D). This figure shows the difference in power in all analyzed frequency bands between the two respective conditions. Time-frequency clusters (defined a priori based upon previous literature) in which the particular mismatch differed from the Correct condition are indicated with a red square. Displayed is the average power difference over all electrodes. Scalp topographies of significant differences between conditions are also displayed. Note the difference in scaling between lower and higher frequencies.

Please cite this article as: Willems, R.M., et al., Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension, Brain Res. (2008), doi:10.1016/j.brainres.2008.04.065

participants' data. However, the power differences at electrode T7 were much less consistent over participants and the effect is mostly due to one outlier in the data. It seems that the effect observed on electrode T7 in the Picture mismatch vs. Correct condition comparison (Fig. 3B) is best explained as an artifact that was not detected in the artifact detection procedure. Note that an ANOVA without the data from electrode T7 yielded similar results: a main effect of Condition ($F(3,45)=3.76$, MSe = 0.30, $p=0.026$) and a significant difference between Picture mismatch and Correct condition ($F(1,15)=4.32$, MSe = 0.323, $p=0.054$) as well as between Picture mismatch and Language mismatch ($F(1,15)=14.00$, MSe = 0.37, $p=0.002$) (Fig. 3D).

In the late gamma cluster (40–50 Hz, 350–750) there was no main effect of Condition ($F(3,45)=1.07$, MSe = 0.27, $p=0.365$) or a Condition × Electrode interaction ($F(78,1140)=1.32$, MSe = 0.22, $p=0.20$).

In summary, first, we observed late increases in theta power (4–6 Hz, 350–750 ms) in all mismatch conditions as compared to the Correct condition. Second, the Double mismatch condition differed significantly from the Correct condition in the early alpha cluster (8–12 Hz, 0–300 ms) and the Language mismatch vs Correct condition showed a trend towards a stronger decrease in the Language mismatch condition. Importantly, a direct comparison between Language mismatch and Picture mismatch revealed that the decrease in alpha power in the early time window is significantly stronger in the Language mismatch condition as compared to the Picture mismatch condition. Finally, we observed a significant decrease in power in the early gamma cluster in the Picture mismatch condition as compared to the Correct condition as well as compared to the Language mismatch condition.

2.2.2. *Analysis without a priori defined time-frequency windows*

The analysis that we employed so far is restricted to the frequency bands in which we expected to find an effect based upon previous literature. To additionally test for time-frequency-electrode clusters that showed differences between conditions, but are not within these pre-defined frequency bands of interest, we employed an analysis which identifies time-frequency-electrode clusters which exhibit consistent changes between conditions. We did this by means of a non-parametric randomization procedure which is described in more detail in the Experimental procedures (see also Maris, 2004; Maris and Oostenveld, 2007). Such analysis has the potential to detect changes between conditions that go unnoticed in the analysis that we employed so far. Given the limited amount of previous research in this field of investigation, such additional effects may very well be present.

A summary of the results of this analysis is provided in Table 4. In general, the results of this additional analysis corroborated the findings from the previous analysis. Two significant increases in power in the 4–6 Hz frequency band were observed. First, the Language mismatch vs. Correct condition led to a significant increase of power in the 4–6 Hz (theta) frequency range from 600 to 750 ms. Second, for the Double mismatch condition vs. Correct condition a similar increase in power in the 4–6 Hz (theta) frequency range from 550 to 750 ms was observed.

A significant decrease in power in the 6–10 Hz (alpha) frequency range was observed between 50 to 200 ms for the Double mismatch vs. Correct condition. A comparable decrease in the Language mismatch vs. Correct condition was not statistically significant ($p=0.13$).

The comparison between the Picture mismatch condition and the Correct condition led to one significant decrease in power in the 40–65 Hz (gamma) frequency range from 0 to 250 ms.

Finally, a direct comparison of the Picture mismatch condition to the Language mismatch condition revealed one significant cluster, involving a decrease in the 40–55 Hz frequency range from 0 to 250 ms to the Picture mismatch condition. No other differences were observed between the experimental conditions (Table 4).

Overall, these results are highly similar as compared to the analysis we performed with a priori defined time-frequency windows of interest. Importantly, no additional clusters of significant changes between conditions were observed.

Table 4 – Results of the time-frequency analysis with the cluster-randomization analysis

Comparison	Time (ms)	Freq (Hz)	Difference	*p*-value	Electrodes
Language mismatch vs. Correct condition	600–750	4–6	Increase	**0.016**	F7, F3, Fz, F4, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, 01, 02
Picture mismatch vs. Correct condition	0–250	40–65	Decrease	**0.028**	F3, Fz, F4, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3 Pz, P4, O1
Double mismatch vs. Correct condition	550–750	4–6	Increase	**0.030**	F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, C3, Cz, C4, T8, CP1, CP2, CP6, P7, P3 Pz, P4, P8, 01, 02
	50–200	6–10	Decrease	**0.043**	F7, F3, Fz, FC5, FC1, FCz, T7, C3, Cz, C4, CP5, CP1, CP2, CP6, P3, Pz, P4, P8
Picture mismatch vs. Language mismatch	0–250	40–55	Decrease	**0.021**	F3, Fz, F4, F8, FC1, FCz, FC2, FC6, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, O1
Double mismatch vs. Language mismatch	–	–	–	–	
Double mismatch vs. Picture mismatch	–	–	–	–	

This analysis identifies consistent differences between conditions over time-frequency-electrode clusters without the need to define time-frequency windows of interest a priori. Displayed are the contrasts that were tested and the significant time window, frequency range and electrodes in which a cluster of increased or decreased activation was detected (see Experimental procedures). 'Decrease' denotes a relative decrease in power, 'increase' denotes a relative increase in power.

Please cite this article as: Willems, R.M., et al., Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension, Brain Res. (2008), doi:10.1016/j.brainres.2008.04.065

As described below, we suspected that the early decreases in alpha/gamma frequency band would be modulated by the amount of prediction which develops over the time course of the experiment. We tested for a linear correlation between item order and effect size in the early (0–300 ms) gamma range (40–50 Hz) between Picture mismatch and the Correct condition, as well as in the early (0–300 ms) window in the alpha frequency band (8–12 Hz) comparing the Language mismatch vs Correct condition and the Double mismatch vs Correct condition. No statistically significant correlations between effect size and item position were observed (all $p>0.25$), indicating that the effect size in the alpha and gamma clusters did not increase or decrease over the course of the experiment. Neither did analyzing the effect sizes in four separate time-bins reveal an effect of item order (see Experimental procedures for details).

3. Discussion

We investigated whether visual and linguistic information presented during language comprehension can be distinguished in terms of changes in power in specific frequency bands of the EEG signal. To increase semantic load of either word and/or picture during sentence comprehension we employed a modified version of the N400 paradigm in which either a word, a picture, or both word and picture could be semantically incongruous with respect to the preceding part of the sentence. Indeed, we found that early decreases in power in the alpha and gamma frequency bands were specifically related to linguistic or visual information load. Before we discuss these specific effects, it should be noted that we replicated earlier findings of similar N400 effects for all three experimental conditions. That is, qualitatively similar N400 effects were observed regardless of the format (linguistic or visual) in which incongruous semantic information was conveyed. For discussion of the theoretical implication of this finding we refer to two recent papers from our lab (i.e. Hagoort and van Berkum, 2007; Willems et al., in press). An increase in theta power was observed for all mismatch conditions. It is tempting to interpret the theta power increase as an oscillatory counterpart of the N400 effect. Analogous to the hypothesized role of the N400, theta increases may reflect increases in semantic integration load. That is, the harder it is to integrate information into a context representation, the stronger the theta power increase. Indeed, conversion of the ERP difference waves into the time-frequency domain, shows an increase around 4 Hz, comparable to the theta increase observed in the frequency analysis (Fig. 4). This pattern suggests that an increase in theta power and the N400 effect may reflect at least partially overlapping cognitive processes. However, previous findings of theta power increases during language comprehension suggest that the role of theta power increases may be broader than just semantic processing. That is, although theta increases have been implicated in semantic processing (Hagoort et al., 2004; Bastiaansen et al., 2005; Hald et al., 2006), they are also reported for increased syntactic processing (Bastiaansen et al., 2002) and working memory operations in general (e.g. Klimesch, 1999). This is not the case for the N400 effect which is closely linked to semantic processing (Kutas and Van Petten, 1994; Brown et al., 2000). A viable possibility is that the theta increases observed here are a reflection of increased (working) memory load, as was suggested recently (Bastiaansen and Hagoort, 2006).

Specifically related to linguistic information, we observed an early (around 150 ms) decrease in alpha power in the Double mismatch condition. Moreover, a direct comparison between Language mismatch and Picture mismatch conditions showed that the decrease in alpha power was significantly stronger in the Language mismatch condition. However, this effect failed to reach significance in the Language mismatch condition vs Correct condition comparison. The alpha band has been claimed to be implicated in general levels of attention or vigilance (Klimesch, 1999), in semantic processing in language tasks (Rohm et al., 2001) and in working memory processes (Jensen et al., 2002; Jokisch and Jensen, 2007). It is however not directly evident how this previous literature can explain the decrease in alpha that we observed. General effects of attention/vigilance or working memory are unlikely to have been different in the Language and Double mismatch conditions compared to the Picture mismatch condition. Second, the occurrence of the effect seems to be too early to reflect a full semantic analysis of the

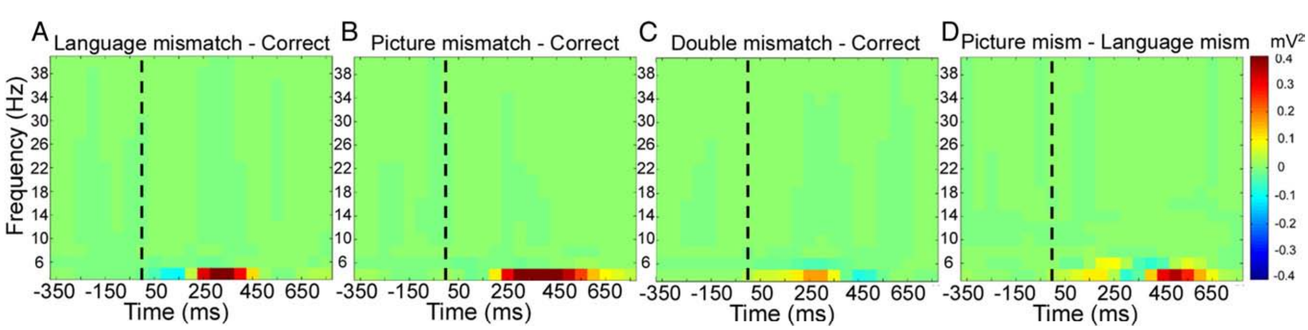

Fig. 4 – Time-frequency representation of the averaged Event-Related Potentials. This figure shows a time-frequency representation of the ERP difference waves between two conditions. Displayed are the TFRs of the difference waves of the Language mismatch-Correct condition (A), Picture mismatch-Correct condition (B), Double mismatch-Correct condition (C) and the Picture mismatch-Language mismatch condition (D). The manifestation of the N400 as an increase in power around 4 Hz is clearly visible. TFRs were created by applying the same analysis procedure for the averaged ERP difference waves as used in the time-frequency analysis of the single trial data.

Please cite this article as: Willems, R.M., et al., Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension, Brain Res. (2008), doi:10.1016/j.brainres.2008.04.065

critical word in connection to the preceding context. An alternative explanation that we entertain here is that a decrease in alpha power reflects an early detection of mismatch in the observed acoustic input based upon the preceding sentence context. That is, in the Language mismatch condition as well as in the Double mismatch condition the incoming acoustic information from the mismatching critical word is different from the 'correct' or matching critical word from the first phoneme on. It is hypothesized that a rapid, context-based analysis of the acoustic information that was not in accordance with the preceding context is at the basis of the alpha power decrease. This effect is reminiscent of the N200 effect that has been reported when the onset of a spoken word differs from the word onset of words that form a congruous completion (Hagoort and Brown, 2000; van den Brink et al., 2001).

Interestingly, Weiss and Rappelsberger (1998) observed a less wide-spread alpha band desynchronisation in reaction to auditory presented words as compared to visually presented words. Similarly, Krause and colleagues observed less strong desynchronisation in the alpha frequency band to auditory presented words as compared to visually presented words (Krause et al., 2006; see also Krause et al., 1997). In the light of the present findings, it is important to note that these studies suggest that hearing language leads to different effects in the alpha frequency band as compared to reading language. Auditory language leads to less alpha band power as compared to visually presented language. This may seem at odds with our present findings of a decrease in alpha band power for a mismatching word. However, it is possible that the detection of a mismatch in acoustic form interferes with the standard processing of an auditory presented word and leads to a power decrease.

Our explanation implies that the upcoming word is at least to some extent predicted. That is, detection of acoustic form which is not in accordance with the previous context, necessitates that another form was expected. Prediction has been shown to play a role in language comprehension and to influence the N400 effect (DeLong et al., 2005; Van Berkum et al., 2005). There is some previous literature which suggests that decrease in alpha band power is sensitive to effects of context on the processing of linguistic stimuli. For instance, Krause et al. (1999) observed a larger decrease to repetition of the same word as compared to the presentation of two different words (see also Karrasch et al., 1998). Klimesch et al. (1990) observed strong effects of expectancy of words and numbers on the amount of alpha band desynchronisation (Klimesch et al., 1990). That is, power decreases in alpha power were stronger when only words were presented as compared to when words and numbers were randomly intermixed (Klimesch et al., 1990). Although this result indicates that decreases in alpha power are related to expectancy of an item, it should be noted that all stimuli in Klimesch et al. were presented visually. That is, the relation to expectancy was not found to be specific for auditory stimuli, as was the case in our study. Finally, it has been found that the processing of a novel, unexpected stimulus leads to less synchronisation in the alpha frequency band between cortical areas in the cat as compared to an expected stimulus (von Stein et al., 2000). Given these previous findings of an influence of expectancy/prediction on alpha power, we considered the effect of prediction in our data. In the present study, the cloze probability of the correct critical words (the amount of participants that filled in the critical word in a pretest, see Experimental procedures) was low (16% on average), so that it is unlikely that expectation was high for the critical words. Moreover, there was no correlation between the effect size of Language mismatch minus Correct condition or of Double mismatch minus Correct condition in the alpha cluster with the order of items. Put differently, the size of the alpha decrease did not increase or decrease over the time course of the experiment. Such an effect might have been expected if participants are able to predict the upcoming critical word when they have become familiar with the set of critical words. However, an alternative explanation is that the language system does predict upcoming words, but that the expectation is not stable over individuals (as reflected in low cloze probabilities). Unfortunately, there is no way in which we can test this assertion in the present study.

An early decrease in the gamma frequency band was observed for the Picture mismatch condition, both in comparison to the Correct condition as compared to the Language mismatch condition. Therefore, this seems to be a specific effect to increased semantic load for information conveyed visually, as through a picture. Analogously to the alpha decrease, we interpret this finding as reflecting an early detection of a mismatch between visual information from the picture and the preceding context. That this effect is manifested in the gamma frequency band is in line with a large body of literature showing successful object recognition to be associated with gamma power increases (see e.g. Rodriguez et al., 1999; Tallon-Baudry, 2003). Our data extend the role of gamma band oscillations in visual object processing to be also related to early detection of a mismatch as compared to a preceding sentence context. That is, these oscillations are not only sensitive to the presentation of an object in isolation, but also to the semantic fit of that object with regard to a preceding (sentence) context. One may argue that the effect occurs too early (around 125 ms) to be a viable candidate of a context-based visual form analysis. Especially the onset of the effect at 0 ms (as determined in the cluster-randomization analysis) is very unlikely. This is however a consequence of the necessary smearing over time as a result of the moving time window used in conversion of the signal into a time-frequency representation. Moreover, early indices (<150 ms) of a rough semantic analysis based on visual form properties have been reported before in the ERP literature (Thorpe et al., 1996). That is, Thorpe and colleagues had subjects classify rich visual scenes (photographs) on the basis whether an animal was present in the scene or not. It was found that within 150 ms after picture onset, the ERP started to diverge based upon the presence or absence of an animal (Thorpe et al., 1996; VanRullen and Thorpe, 2001). Related to this, it has been shown that the visual context in which a face is presented modulates the ERP response 170 ms after stimulus onset (Righart and de Gelder, 2006). Summarized, there is evidence from previous ERP literature that recognition of objects in complex scenes as well as effects of context can have an effect upon the ERP response within or around 150 ms after stimulus presentation.

Please cite this article as: Willems, R.M., et al., Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension, Brain Res. (2008), doi:10.1016/j.brainres.2008.04.065

Interestingly, a recent study reports a similarly early gamma decrease to an incongruous *word* (no pictures were presented) (Hald et al., 2006). Stimuli were written sentences, presented word-by-word, as opposed to the spoken stimulus materials in the present study. In analogy to our explanation of the present gamma decrease, it is possible that an early, context-based visual analysis also occurs in the case of a visual word form. That is, in the case of written language, an early context-based detection of visual word form mismatch may have caused the early decrease in gamma band power in the study by Hald et al. Such an explanation is indirectly supported by the fact that an early decrease in alpha band power that we linked to acoustic mismatch (above), was not observed in the Hald et al. study.

One reason to be cautious about the interpretation of the gamma decrease is that such decrease is not observed in the Double mismatch condition. If the early gamma decrease in the Picture mismatch condition reflects the context-based detection of an early mismatch of sentence context and visual form features of the picture, it is unclear why this effect is absent in the Double mismatch condition. After all, also in the Double mismatch condition the picture is not in accordance with the preceding sentence context. A different explanation is that instead of being a picture-specific effect, the gamma decrease reflects an early detection of conflicting information coming from co-occurring critical word and picture regardless of the previous context. Such an explanation is supported by the recent finding that incongruous sound and picture of an animal leads to reduced gamma frequency band activation as compared to matching sounds and pictures (Yuval-Greenberg and Deouell, 2007). If this were the case, we would also expect a gamma decrease in the Language mismatch condition, in which co-occurring picture and word also convey different information. This was not observed however.

A recent hypothesis links gamma power increases to a successful match between stored representations in memory and incoming stimulus information (Herrmann et al., 2004b). For instance, Herrmann and colleagues observed a gamma (30–40 Hz) increase in response to pictures of existing objects as compared to nonsense objects, 90 ms after stimulus presentation (Herrmann et al., 2004a). From this it was concluded that the successful matching of the stimulus picture with the presence of a representation of the object in long-term memory is reflected in increased gamma power over occipital electrodes. Along these lines, it may be the case that in the present study the sentence context primed expectation of a certain picture (cf. Engel et al., 2001). In the case of the Picture mismatch condition, this expectation is violated which leads to a lack of increase in gamma power. This is in line with our explanation for the decrease in gamma power in the Picture mismatch condition. The fact that we did not observe differences in early gamma power to a mismatching word in the Language mismatch condition however seems to contradict the hypothesis that early gamma band increase is a reflection of matching of incoming stimulus information with a long-term representation *independent of input format* (Herrmann et al., 2004b; Lenz et al., 2007). Given our findings, this model could be extended in the sense that it seems that the format of the incoming information makes a crucial difference. After all, no effects in the gamma band were observed in the Language mismatch condition.

In conclusion, we hypothesize that the late increases in theta band power reflect general working memory processes related to the integration of semantic information into a representation of the sentence context. An early decrease in alpha band power most likely reflects early analysis of the acoustic input which was not in line with the acoustic input as expected from the preceding sentence context. Finally, an early decrease in gamma band power is tentatively explained as reflecting detection of a mismatch of visual form features on the basis of the preceding context. With regard to our main question we conclude that semantic integration during sentence comprehension is neurally implemented in a similar way regardless of the format the input is in. Early differences in specific frequency bands however code for differences in the way the incoming information is conveyed. Crucially, these early differences are context-dependent in that they only occur when the item is semantically incongruous with the preceding (sentence) context. As such they are hypothesized to reflect an early, coarse semantic analysis based upon the acoustic or visual form of the input.

As a final note we want to point out that this study is one in a series in the neurocognition of language that shows the value of doing complementary analyses in the time-frequency domain next to more traditional ERP analysis. Qualitative differences in neural processing between visual and linguistic information that could not be detected in the grand average ERPs, were observed in induced changes in specific frequency bands. However, we acknowledge that firm interpretation of the results is hindered by the lack of a considerable amount of studies in this field of research. It will be a challenge for future EEG research on language to employ frequency analysis to find a tighter explanation of the role of oscillations during language comprehension.

4. Experimental procedures

4.1. *Stimulus materials*

A total of 328 sentences (mean duration 3196 ms, range 2164–4184 ms) were recorded in a sound attenuated room at 44.1 KHz, spoken at a normal rate by a native Dutch female speaker. Half of these sentences differed in one critical word, which was never in sentence final position. In each sentence a short context was introduced to which the critical word was congruous or not. Critical words were nouns that corresponded to names given by a separate group of participants ($n=32$) to a large set of black and white line drawings. All critical words had a picture equivalent with a naming consistency of 85% or higher. In total there were 26 critical words with their picture equivalents. All words were one syllable long and started with a plosive consonant. Every critical word occurred equally often in a matching and in a mismatching sentence context. The critical word in the mismatching sentence always had a different onset consonant than the critical word in the semantically correct sentence. Sentences were pretested in a cloze probability test by a separate participant group ($n=16$). The percentage of participants that gave the target word as response was taken as a measure of its cloze probability. Overall, the mean cloze probability was 16% for the semantically congruous critical words (range 0–69%), and 0% for the semantically incongruous critical words.

4.2. Participants

Data of three participants in our original data set (Willems et al., in press) had to be discarded because of excessive (muscle) artifacts in high frequency bands. These were replaced by three novel data sets, such that data of 16 participants went into the analysis (mean age = 22.8 years, range 18–34, 13 females). All participants were healthy, right-handed (Oldfield, 1971), and had Dutch as their mother tongue. None had any known neurological history, hearing complaints and all had normal or corrected-to-normal vision. Participants were paid for participation. The local ethics committee approved the study and all participants signed informed consent in accordance with the declaration of Helsinki.

4.3. Procedure

Stimuli were presented using Presentation software (version 9.13, http://www.neurobs.com/). Four stimulus lists of 164 trials each were created in which only one item of every stimulus quartet (as in Table 1) was presented. Sentences were pseudo-randomized with the constraint that the same condition occurred maximally two times in a row. Every list contained an equal amount of stimuli from the four conditions (41 per condition). Pictures had varying sizes depending upon the object they represented and were maximally 8 × 8 cm (5° × 5° visual angle; minimum height × width: 2.5 cm × 7.5 cm and 7 cm × 3 cm), shown at a viewing distance of 90 cm. Pictures were presented from the onset of the critical word to the end of the sentence. A trial started with 600 ms blank screen, followed by a spoken sentence and a picture, 1000 ms blank screen and 2500 ms with a fixation cross on the screen. Participants were instructed to sit still in a comfortable position and to blink only when the fixation cross was presented. The session started with eight practice trials which contained different critical words than used in the main part of the experiment. Participants were told to attentively listen to and watch the stimuli about which they would receive questions afterwards. At the end of the test session, general questions about the stimuli were asked. All participants had understood the manipulation in the materials and could provide examples of stimuli.

4.4. Recording

The electroencephalogram (EEG) was recorded from 27 electrode sites across the scalp using an Electrocap with Ag/AgCl electrodes, each referred to the left mastoid. Electrodes were placed on standard electrode sites (Fz, FCz, Cz, Pz, F3, F4, F8, F7, FP2, FC5, FC1, FC2, FC6, T7, T8, C3, C4, CP5, CP1, CP2, CP6, P7, P3, P4, P8, O1, O2). Vertical eye movements and blinks were monitored by means of two electrodes, one placed beneath and one above the left eye. Horizontal eye movements were monitored by means of a left to right bicanthal montage. Activity over the right mastoid was recorded to determine if there were additional contributions of the experimental variables to the two presumably neutral mastoid sites. No such differences were observed. Recordings were amplified with BrainAmp DC amplifiers, using a hi-cut of 100 Hz and a time constant of 10 s. Impedances were kept below 5 kΩ for all channels. The EEG and EOG signals were recorded and digitized using Brain Vision Recorder software (version 1.03), with a sampling frequency of 500 Hz.

4.5. Analysis

Analyses were done using the FieldTrip software package, which is an open-source Matlab toolbox designed for EEG/MEG data analysis (http://www.ru.nl/fcdonders/fieldtrip/). Data were filtered off-line with a 70 Hz low-pass filter, re-referenced to the mean of the two mastoids and segmented from 600 ms before to 1000 ms after the critical word. All segments were screened for eye movements, electrode drifting, amplifier blocking and muscle artifacts, leading to 32% of the trials to be rejected, equally distributed over conditions ($F < 1$), which is comparable to the amount of trials analyzed in other studies investigating oscillatory correlates of sentence comprehension (Hagoort et al., 2004; Hald et al., 2006; Davidson and Indefrey, 2007).

For the ERP analysis, baseline correction was applied by subtracting the mean of the pre-stimulus period from 150 ms to the onset of the critical word. ERPs were created by averaging all trial segments for each condition and subject separately. Statistical analysis was performed by employing repeated measures analysis of variance (ANOVA) on the mean amplitude in the 300–600 ms time window with factors Condition (4 levels; Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 levels; Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2). In the case of a main effect of Condition or Condition × Electrode interaction, subsequently, planned comparisons were performed to test for differences between each mismatch condition and the Correct condition as well as between the Language mismatch condition and the Picture mismatch condition. Huynh–Feldt correction for violation of sphericity assumption was applied when appropriate (Huynh and Feldt, 1976), but original degrees of freedom are reported (Table 2).

For the time-frequency analysis, the time-frequency representation (TFR) was computed for every trial using a multitaper procedure (Mitra and Pesaran, 1999). Low (4–40 Hz) and high frequencies (40–70 Hz) were analyzed separately. For the low frequencies a 500 ms sliding window with a single Hanning taper was used with no spectral smoothing. For the high frequencies, a 200 ms sliding window with three orthogonal Slepian tapers was used with 10 Hz spectral smoothing. Note that conversion into the frequency domain limited the maximal time-point of a segment which could be estimated at 750 ms (low frequencies) and 900 ms (high frequencies) after critical word onset. In the analysis, each segment was analyzed up to 750 ms after critical word onset. Average TFRs were computed by averaging single trial TFRs for every condition and subject separately.

Subsequently, statistical analysis involved repeated measures analysis of variance (ANOVA) on mean power in six a priori defined time-frequency windows with factors Condition (4 levels; Correct condition, Language mismatch, Picture mismatch, Double mismatch) and Electrode (27 levels; Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FCz, FC2, FC6, T7, C3, Cz, C4, T8, CP5, CP1, CP2, CP6, P7, P3, Pz, P4, P8, O1, O2). The a priori defined frequency windows were based upon previous literature and included the theta band (4–6 Hz) (Bastiaansen et al., 2005; Hald et al., 2006; Davidson and Indefrey, 2007), alpha band (8–12 Hz) (Rohm et al., 2001) and the lower gamma band (40–50 Hz) (Hagoort et al., 2004; Hald et al., 2006). Separate ANOVAs were conducted in early (0–300 ms) and late time windows (350–750 ms). In the case of a main effect of Condition or Condition × Electrode interaction, subsequently, planned comparisons were performed to test for differences between each mismatch condition and the Correct condition as well as between the Language mismatch condition and the Picture mismatch condition. Huynh–Feldt correction for violation of sphericity assumption was applied when appropriate (Huynh and Feldt, 1976), but original degrees of freedom are reported.

Moreover, to test for significant differences between conditions in time-frequency-electrode clusters that were outside of the a priori defined time-frequency windows, the data were analyzed using a cluster-randomization procedure which identifies consistent changes between conditions in time-frequency-electrode clusters. First, single-subject statistics were computed (two-sided *t*-test for the difference between two conditions for every electrode-time-frequency point). Consequently, group statistics involved a clustering procedure on the tresholded ($t>1.96$ and $t<-1.96$) single-subject statistics which identifies clusters of time-frequency-electrode points showing the same direction of effect (Maris, 2004). To assess statistical significance of each cluster, the sum of all t-statistics in the cluster was computed. This was chosen as the cluster-level statistic (Maris and Oostenveld, 2007). Second, significance of each cluster-level statistic was assessed by comparing the cluster statistic to its randomization distribution which was created by 2500 random re-assignments of the single-subject statistics and zero. That is, in each randomization the single-subject statistics were randomly re-assigned to zero or to their original value. The actual cluster statistic was then compared to the randomization distribution obtained and significance was assessed by evaluating the cluster statistic to the $p<0.05$ significance level. Note that the validity of the inference drawn is not dependent upon the exact statistic (in this case sum of all t-statistics) chosen (as is explained in more detail in Maris and Oostenveld, 2007). The choice for the sum of individual t-statistics is based upon theoretical/modeling considerations (Maris, 2004; Maris and Oostenveld, 2007) as well as by practice in other studies employing a cluster-randomization procedure (e.g. Hald et al., 2006; Osipova et al., 2006; Davidson and Indefrey, 2007; Medendorp et al., 2007; Tuladhar et al., 2007). This non-parametric procedure effectively controls the multiple comparisons problem introduced by the massive univariate approach taken (Maris and Oostenveld, 2007).

We tested the hypothesis that some effects may have decreased/increased over the time course of the experiment (see Discussion section) in two ways. First, it was tested whether there was a linear correlation between effect size and item order in three time-frequency windows in which such correlation may have been expected (early time window (0–300 ms) in alpha frequency band (8–12 Hz) in Language mismatch–Correct condition and Double mismatch–Correct condition comparisons and early time window (0–300 ms) in the gamma frequency band (40–50 Hz) in the Picture mismatch–Correct condition comparison). Second, data from each of these three time-frequency windows were split into four equal parts according to their occurrence in the experiment and subjected to repeated measures analysis of variance (ANOVA) with factors Time (4 levels) and Electrode (27 levels). No effects of item order/time course of the experiment were found.

Acknowledgments

This research was supported by a grant from the Netherlands Organization for Scientific Research (NWO 'Cognition' program, grant no 051.02.040). Petra van Alphen is acknowledged for expertly voicing the sentences. We thank Marcel Bastiaansen for the valuable comments on an earlier version of our manuscript, Heidi Koppenhagen and Niels Schiller for providing the naming consistency information of the line drawings and two anonymous reviewers for their valuable comments on our manuscript.

REFERENCES

Bastiaansen, M.C., Hagoort, P., 2006. Oscillatory neuronal dynamics during language comprehension. Prog. Brain Res. 159, 179–196.

Bastiaansen, M.C., van Berkum, J.J., Hagoort, P., 2002. Syntactic processing modulates the theta rhythm of the human EEG. NeuroImage 17 (3), 1479–1492.

Bastiaansen, M.C., van der Linden, M., Ter Keurs, M., Dijkstra, T., Hagoort, P., 2005. Theta responses are involved in lexical-semantic retrieval during language processing. J. Cogn. Neurosci. 17 (3), 530–541.

Brown, C.M., Hagoort, P., Kutas, M., 2000. In: Gazzaniga, M.S. (Ed.), Postlexical integration processes in language comprehension: Evidence from brain-imaging research. The cognitive neurosciences. MIT Press, Cambridge, Mass, pp. 881–895.

Davidson, D.J., Indefrey, P., 2007. An inverse relation between event-related and time-frequency violation responses in sentence processing. Brain Res. 1158, 81–92.

DeLong, K.A., Urbach, T.P., Kutas, M., 2005. Probabilistic word pre-activation during language comprehension inferred from electrical brain activity. Nat. Neurosci. 8 (8), 1117–1121.

Engel, A.K., Fries, P., Singer, W., 2001. Dynamic predictions: oscillations and synchrony in top-down processing. Nat. Rev. Neurosci. 2 (10), 704–716.

Federmeier, K.D., Kutas, M., 2001. Meaning and modality: influences of context, semantic memory organization, and perceptual predictability on picture processing. J. Exp. Psychol.: Learn. Mem. Cogn. 27 (1), 202–224.

Ganis, G., Kutas, M., 2003. An electrophysiological study of scene effects on object identification. Cogn. brain res. 16 (2), 123–144.

Ganis, G., Kutas, M., Sereno, M.I., 1996. The search for "common sense": an electrophysiological study of the comprehension of words and pictures in reading. J. Cogn. Neurosci. 8 (2), 89–106.

Hagoort, P., Brown, C.M., 2000. ERP effects of listening to speech: semantic ERP effects. Neuropsychologia 38 (11), 1518–1530.

Hagoort, P., van Berkum, J., 2007. Beyond the sentence given. Philos. Trans. R. Soc. Lond., B Biol. Sci. 362 (1481), 801–811.

Hagoort, P., Hald, L., Bastiaansen, M., Petersson, K.M., 2004. Integration of word meaning and world knowledge in language comprehension. Science 304 (5669), 438–441.

Hald, L.A., Bastiaansen, M.C., Hagoort, P., 2006. EEG theta and gamma responses to semantic violations in online sentence processing. Brain Lang. 96 (1), 90–105.

Herrmann, C.S., Lenz, D., Junge, S., Busch, N.A., Maess, B., 2004a. Memory-matches evoke human gamma-responses. BMC Neurosci. 5, 13.

Herrmann, C.S., Munk, M.H., Engel, A.K., 2004b. Cognitive functions of gamma-band activity: memory match and utilization. Trends Cogn. Sci. 8 (8), 347–355.

Huynh, H., Feldt, L., 1976. Estimation of the Box correction for degrees of freedom from sample data in randomized block and splitplot designs. J. Educ. Stat. 1, 69–82.

Jensen, O., Gelfand, J., Kounios, J., Lisman, J.E., 2002. Oscillations in the alpha band (9–12 Hz) increase with memory load during retention in a short-term memory task. Cereb. Cortex 12 (8), 877–882.

Jensen, O., Kaiser, J., Lachaux, J.P., 2007. Human gamma-frequency oscillations associated with attention and memory. Trends Neurosci. 30 (7), 317–324.

Jokisch, D., Jensen, O., 2007. Modulation of gamma and alpha activity during a working memory task engaging the dorsal or ventral stream. J. Neurosci. 27 (12), 3244–3251.

Karrasch, M., Krause, C.M., Laine, M., Lang, A.H., Lehto, M., 1998. Event-related desynchronization and synchronization during an auditory lexical matching task. Electroencephalogr. Clin. Neurophysiol. 107 (2), 112–121.

Klimesch, W., 1999. EEG alpha and theta oscillations reflect cognitive and memory performance: a review and analysis. Brain Res. Brain Res. Rev. 29 (2–3), 169–195.

Klimesch, W., Pfurtscheller, G., Mohl, W., Schimke, H., 1990. Event-related desynchronization, ERD-mapping and hemispheric differences for words and numbers. Int. J. Psychophysiol. 8 (3), 297–308.

Krause, C.M., Porn, B., Lang, A.H., Laine, M., 1997. Relative alpha desynchronization and synchronization during speech perception. Brain Res. Cogn. Brain Res. 5 (4), 295–299.

Krause, C.M., Astrom, T., Karrasch, M., Laine, M., Sillanmaki, L., 1999. Cortical activation related to auditory semantic matching of concrete versus abstract words. Clin. Neurophysiol. 110 (8), 1371–1377.

Krause, C.M., Gronholm, P., Leinonen, A., Laine, M., Sakkinen, A.L., Soderholm, C., 2006. Modality matters: the effects of stimulus modality on the 4-to 30-Hz brain electric oscillations during a lexical decision task. Brain Res. 1110 (1), 182–192.

Kutas, M., Hillyard, S.A., 1980. Reading senseless sentences: brain potentials reflect semantic incongruity. Science 207 (4427), 203–205.

Kutas, M., Van Petten, C.K., 1994. In: Gernsbacher, M.A. (Ed.), Psycholinguistics electrified: Event-related brain potential investigations. Handbook of psycholinguistics. Academic Press, San Diego, CA, pp. 83–143.

Lenz, D., Schadow, J., Thaerig, S., Busch, N.A., Herrmann, C.S., 2007. What's that sound? Matches with auditory long-term memory induce gamma activity in human EEG. Int. J. Psychophysiol. 64 (1), 31–38.

Maris, E., 2004. Randomization tests for ERP topographies and whole spatiotemporal data matrices. Psychophysiology 41 (1), 142–151.

Maris, E., Oostenveld, R., 2007. Nonparametric statistical testing of EEG- and MEG-data. J. Neurosci. Methods 164 (1), 177–190.

Medendorp, W.P., Kramer, G.F., Jensen, O., Oostenveld, R., Schoffelen, J.M., Fries, P., 2007. Oscillatory activity in human parietal and occipital cortex shows hemispheric lateralization and memory effects in a delayed double-step saccade task. Cereb. Cortex 17 (10), 2364–2374.

Mitra, P.P., Pesaran, B., 1999. Analysis of dynamic brain imaging data. Biophys. j. 76, 691–708.

Nigam, A., Hoffman, J.E., Simons, R.F., 1992. N400 to semantically anomalous pictures and words. J. Cogn. Neurosci. 4 (1), 15–22.

Oldfield, R.C., 1971. The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9 (1), 97–113.

Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., Jensen, O., 2006. Theta and gamma oscillations predict encoding and retrieval of declarative memory. J. Neurosci. 26 (28), 7523–7531.

Özyürek, A., Willems, R.M., Kita, S., Hagoort, P., 2007. On-line integration of semantic information from speech and gesture: Insights from event-related brain potentials. J. Cogn. Neurosci. 19 (4), 605–616.

Righart, R., de Gelder, B., 2006. Context influences early perceptual analysis of faces-an electrophysiological study. Cereb. Cortex 16 (9), 1249–1257.

Rodriguez, E., George, N., Lachaux, J.P., Martinerie, J., Renault, B., Varela, F.J., 1999. Perception's shadow: long-distance synchronization of human brain activity. Nature 397 (6718), 430–433.

Rohm, D., Klimesch, W., Haider, H., Doppelmayr, M., 2001. The role of theta and alpha oscillations for language comprehension in the human electroencephalogram. Neurosci. Lett. 310 (2–3), 137–140.

Sitnikova, T., Kuperberg, G., Holcomb, P.J., 2003. Semantic integration in videos of real-world events: an electrophysiological investigation. Psychophysiology 40 (1), 160–164.

Tallon-Baudry, C., 2003. Oscillatory synchrony and human visual cognition. J. Physiol. Paris 97 (2–3), 355–363.

Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. Nature 381 (6582), 520–522.

Tuladhar, A.M., ter Huurne, N., Schoffelen, J.M., Maris, E., Oostenveld, R., Jensen, O., 2007. Parieto-occipital sources account for the increase in alpha activity with working memory load. Hum. Brain Mapp. 28 (8), 785–792.

Van Berkum, J.J.A., Brown, C.M., Zwitserlood, P., Kooijman, V., Hagoort, P., 2005. Anticipating upcoming words in discourse: evidence from ERPs and reading times. J. Exp. Psychol.: Learn. Mem. Cogn. 31 (3), 443–467.

van den Brink, D., Brown, C.M., Hagoort, P., 2001. Electrophysiological evidence for early contextual influences during spoken-word recognition: N200 versus N400 effects. J. Cogn. Neurosci. 13 (7), 967–985.

VanRullen, R., Thorpe, S.J., 2001. The time course of visual processing: from early perception to decision-making. J. Cogn. Neurosci. 13 (4), 454–461.

von Stein, A., Chiang, C., Konig, P., 2000. Top-down processing mediated by interareal synchronization. Proc. Natl. Acad. Sci. U. S. A. 97 (26), 14748–14753.

Weiss, S., Rappelsberger, P., 1998. Left frontal EEG coherence reflects modality independent language processes. Brain Topogr. 11 (1), 33–42.

West, W.C., Holcomb, P.J., 2002. Event-related potentials during discourse-level semantic integration of complex pictures. Cogn. brain res. 13 (3), 363–375.

Willems, R.M., Özyürek, A., Hagoort, P., 2007. When language meets action: the neural integration of gesture and speech. Cereb. Cortex 17 (10), 2322–2333.

Willems, R.M., Özyürek, A., Hagoort, P., in press. Seeing and hearing meaning: ERP and fMRI evidence of word versus picture integration into a sentence context. J. Cogn. Neurosci. doi:10.1162/jocn.2008.20085.

Yuval-Greenberg, S., Deouell, L.Y., 2007. What you see is not (always) what you hear: induced gamma band responses reflect cross-modal interactions in familiar object recognition. J. Neurosci. 27 (5), 1090–1096.

Please cite this article as: Willems, R.M., et al., Early decreases in alpha and gamma band power distinguish linguistic from visual information during spoken sentence comprehension, Brain Res. (2008), doi:10.1016/j.brainres.2008.04.065