Detrimental Effects of Irrelevant Speech on Serial Recall of Visual Items are Reflected in Reduced Visual N1 and Reduced Theta Activity

The term irrelevant sound effect (ISE) describes an empirically robust finding in which serial recall performance of visual items is reduced by irrelevant speech. At present little is known about its neurophysiological basis. Although some previous neuroelectric studies have concentrated on responses elicited by irrelevant background sound, whether the processing of visually presented to-be-remembered digits itself is affected by irrelevant speech has yet to be studied. An experiment (n = 20) was conducted in which serial recall performance for visually presented digits was tested during exposure to either irrelevant speech, continuous white noise or silence while measuring EEG activity. White noise was chosen as a control condition, because it constitutes auditory stimulation while leaving serial recall performance unimpaired. In addition to replicating the detrimental behavioural effect of irrelevant speech, an analogous speech-specific early decrease (~160 ms) in the visual ERP (N1) and subsequently a reduced theta response (4–6 Hz; 400–800 ms) at right prefrontal electrodes were observed. Although irrelevant sound presentation was restricted to the visual presentation phase, power spectra reveal that the weaker theta response for speech persisted in a silent retention phase before serial recall. Based on such data we propose reevaluating the role attention plays in explaining the ISE.

Keywords: EEG, irrelevant sound effect, ISE, N1, theta

Introduction

The Irrelevant Sound Effect

A robust finding in working memory research is that immediate serial recall performance for visually presented digits or letters is reduced significantly by simultaneously presenting irrelevant sounds characterized by distinct temporal-spectral variations (e.g., narration, music with prominent staccato passages, sequences of different sine-wave tones). The so-called irrelevant sound effect (ISE) is therefore defined as the difference between serial recall performance in a silent condition and an irrelevant sound condition. The effect occurs although background sounds are not related to the task (i.e., irrelevant sounds) and subjects are told to ignore them. The present study intends to explore the impact of irrelevant background speech on a cortical level by means of electroencephalography (EEG) and aims to contrast its neurophysiological effects with those of continuous noise, which does not reduce serial recall performance.

The ISE was first reported by Colle and Welsh (1976), and since that time many behavioural experiments unequivocally corroborated the original finding that irrelevant narrative speech detrimentally influences serial recall performance (e.g., Buchner et al., 1996; Salamé and Baddeley, 1986, 1987). However, an ISE can be caused by non-speech sounds, too, e.g. by music (e.g. Morris et al., 1989; Nittono, 1997; Ellermeier and Hellbrück, 1998) or by sequences of different tones (e.g. Jones and Macken, 1993; Jones et al., 1999; Macken et al., 2003). In order to elicit an ISE, it appears to be crucial that the irrelevant sound comprises of distinct auditory-perceptual units which vary consecutively (changing-state characteristic). Irrelevant sounds with little or no temporal-spectral variations (steady-state sounds) typically do not disrupt serial recall performance (e.g. continuous noise; Salamé and Baddeley, 1987; Jones et al., 1990; Ellermeier and Zimmer, 1997).

Current Models Explaining the ISE

The standard task to measure verbal short-term memory capacity is the immediate serial recall of verbal material. The first short-term memory model incorporating the ISE was the working memory model of Baddeley (1986, 1997, 2000). According to this, the detrimental effect irrelevant speech has on serial recall is exclusively located in the verbal subsystem called the phonological loop. Visually presented items are fed into the phonological loop by means of a subvocal rehearsal process. Heard speech, such as irrelevant speech, is assumed to have obligatory access to the loop (e.g. Salamé and Baddeley, 1982; Baddeley et al., 1984; Baddeley and Salamé, 1986), where it putatively interferes with the representations of the to-be-remembered items. However, the interference mechanisms still remain unspecified.

This lack of specification encouraged the development of alternative models, such as the object-oriented episodic record model (O-OER model) of Jones and co-authors (Jones, 1993; Jones et al., 1996; Macken et al., 1999). In this model, all sensory input is deposited directly into one unitary store. The to-be-remembered items are encoded as objects there and are aligned by order information necessary for serial recall. Concurrent order information is automatically set up among the encoded objects of an irrelevant changing-state sound, too. The coexisting order information — on the one hand for the to-be-remembered item-sequence and on the other for the irrelevant changing-state sound — results in its loss and, consequently, an ISE is observed. In contrast, an irrelevant sound with steady-state characteristics (e.g. continuous noise) is encoded as a single object having only self-referential order information. In this case serial-recall performance for visual items remains unaffected by irrelevant sound.

What both models have in common is that they exclude attentional aspects possibly playing a role in ISE evocation or magnitude. Although some findings do exist which favour the ISEs independence of attention, recent neurophysiological and behavioural studies challenge this view. Pros and cons for independence of attention will be presented in the following.

Nathan Weisz1 and Sabine J. Schlittmeier2

1Department of Psychology, University of Konstanz, Germany and 2Department of Psychology, Catholic University of Eichstätt-Ingolstadt, Germany

© The Author 2005. Published by Oxford University Press. All rights reserved.
For permissions, please e-mail: journals.permissions@oxfordjournals.org
In favour of an attentional independency of the ISE is that habituation — as defined by Öhman (1979) — does not occur with irrelevant speech, as shown by behavioural data (Hellbrück et al., 1996; Ellermeier and Zimmer, 1997). Moreover, the detrimental effect irrelevant changing-state sounds exhibit compared to steady-state sounds persists even if successive trials are learned during different sound conditions (e.g. Jones et al., 1997; Tremblay and Jones, 1998). In this case, the auditory background significantly changes at the beginning of each trial thereby causing an orienting reflex as well as an attentional distraction (Sokolov, 1963).

Contrarily, neurophysiological experiments exploring the ISE argue for the involvement of attentional processes. The N1 is known to be sensitive to manipulations of selective attention (Hillyard and Anllo-Vento, 1998; Hillyard et al., 1998) and research shows that it is affected by irrelevant sound. In a visual serial recall task during sequences of different pure tones, Valtonen et al. (2003) found an enhanced auditory neuro-magnetic N1 (elicited by pure tones) for increasing memory load compared to a listen-only condition. An enlarged auditory N1 along with decreased recall performance during irrelevant speech was also found by Campbell et al. (2003) if the number of different meaningless syllables the irrelevant sound consisted of rose from two to five. Recent behavioural studies suggest attention can play a role in generating the ISE as well. For example, the affective valence of words composing irrelevant speech or word frequency contributes to ISE magnitude (Elliott, 2002; Buchner et al., 2004; Buchner and Erdfelder, 2005). This solely appears explainable in terms of attention catching or distraction.

The only explicit theoretical explanation of the ISE which ascribes attention a role is offered by Neath (1999, 2000), explaining the ISE in terms of the feature model (Nairne, 1988, 1990; Neath and Nairne, 1995). Information is represented in this model by vectors of feature values which are built up in primary and secondary memory, a distinction made following William James (1890). The feature vectors must be sampled and matched again for recall. The probability for correct matching and hence serial recall performance are reduced in the presence of irrelevant changing-state sound which is assumed to reduce overall available attentional resources. An ISE is not expected for continuous noise, since it can be easily ignored. While irrelevant non-speech sounds are assumed to only have the mentioned attentional effect (Neath, 2000; Neath and Surprenant, 2001), the feature model additionally assumes irrelevant speech to degrade modality-independent features of the feature vectors encoded in primary memory (feature adoption). This also reduces the probability for correct matching.

Research Interest
Do attentional aspects contribute to the elicitation of an ISE as suggested by neurophysiological and behavioural data as well as by the feature model? Although only few in number, some EEG experiments exploring the ISE have been conducted. Yet most of them focussed on the processing of the irrelevant sound itself (Campbell et al., 2003; Valtonen et al., 2003). No EEG-study hitherto has investigated how the presence of irrelevant sound affects the neural processing of the to-be-remembered items themselves. This is of great importance, however, since a withdrawal of attentional resources of the to-be-remembered items taking place under irrelevant sound can only be demonstrated by showing that this sound condition reduces neuronal activity related to the processing of the memory items. Our experiment focuses on this neglected aspect.

We conducted a typical ISE-experiment in which visual items were presented either during irrelevant speech, white noise or silence, and they had to be recalled in strict serial order. Consistent with research literature, we expected reduced serial recall performance during irrelevant speech but not during continuous noise on a behavioural level. Yet apart from this, our main research question concerned the impact of different sound conditions on the cortical processing of the visual items. Event-related potentials (ERPs) for visual memory items were measured to explore whether irrelevant speech has any specific impact on visual item processing, which continuous noise does not have. Both main ISE-theories proposed by Baddeley and Jones do not ascribe a role to attentional mechanisms for elicitation of the ISE. Given this perspective, Jones and Baddeley do not expect irrelevant speech to have any influence on ERP components known to be sensitive to attentional modulation, such as the P1 and N1 (Hillyard and Anllo-Vento, 1998; Hillyard et al., 1998). In contrast to this, Neath attributes a general attention-demanding effect to irrelevant sound. However, an amplitude reduction (signifying less directed attention towards the to-be-remembered item) is hardly explainable within the feature model either, because modality-dependent feature information is explicitly assumed not to be affected by irrelevant sound.

In addition to the ERPs, a time-frequency (wavelet) analysis was performed in order to capture the effects of irrelevant speech on neural processing which may not be strictly phase-locked across trials. In investigating the silent retention period separating item presentation and recall, we took a glance at the power spectra. Does irrelevant speech have any specific effect on oscillatory EEG measures compared to continuous noise, as oscillatory EEG measures are abundantly reported to be associated with short-term memory? This link is best established for theta oscillations, since they are involved in several short-term memory studies (Raghavachari et al., 2001; Lee et al., 2005) and are known to vary with memory load (Klimesch, 1996; Jensen and Tesche, 2002). If theta waves can be viewed as the brain’s mark for retaining visual information during short-term memory tasks, we predict that reduced theta activation should be singularly observed in our experiment for the irrelevant speech condition.

Materials and Methods

Participants
Twenty right-handed subjects (10 male; age 20–39 years) participated in the study. They were unfamiliar with its specific design and hypotheses. Written informed consent was collected from each individual. All subjects reported normal hearing and normal or corrected-to-normal vision.

Stimuli and Procedure
At the beginning of the experiment, each participant was seated and electrodes were mounted on an EEG-cap (see next section). The distance between the subjects’ head and the monitor was approximately 70 cm.

A serial-recall trial encompassed the successive presentation of the visual to-be-remembered items, a retention interval of 10 s and a recall phase. Note that a 10 s retention interval is not unusual in ISE experiments and has been used in several studies (e.g. Jones et al., 1993; Klatte et al., 2002; Larsen and Baddeley, 2003).
Each trial started with a blank screen being shown for 3020 ms followed by the presentation of a warning stimulus (1500 ms), which consisted of three consecutively presented rectangles (rate: 2 Hz) decreasing in size. The last rectangle was not cleared. After an ISI randomly varied between 800 and 1000 ms (eight ISIs in 25 ms steps), the first digit (in 112 pt Chicago font) was presented in the rectangle located in the middle of the screen for a total of 700 ms, followed again by an ISI varying between 800 and 1000 ms. Presentation durations and ISIs remained the same for the 2-9 digits. In each trial, digits 1-9 were presented in randomized order and taken from a list of 500 permutations excluding non-trivial sequences (excluding e.g. 1, 2, 3, or 4). After the to-be-remembered digits were presented, a 10 s silent retention interval (blank screen) followed. Subsequently, a 3 × 3 display of rectangles appeared in which the digits were displayed. In each trial, the digits were ordered in a random manner to avoid spatial recall strategies. Each participant was requested to click the digits in the sequence they were previously presented. After clicking, the rectangle disappeared. Correcting errors was not possible.

There were three sound conditions depending on whether irrelevant speech (amplitude normalized to the RMS level of the white noise), white noise (power uniformly distributed between 0 and 22.05 kHz; 65 dB) or no sound was played during digit presentation. The sounds (sampled at 44.1 kHz; 16 bit D/A conversion; integrated sound card: Texas Instruments© TAS3004) were diotically presented via headphones (Sennheiser HD280pro), commencing with the onset of the second warning stimulus and ending at the offset of the ninth digit. Irrelevant background conditions were varied quasi-randomly from trial to trial so that no two successive trials involved the same sound condition. Narration by a Japanese speaker was used as irrelevant background speech and was incomprehensible for our non-Japanese-speaking subjects. For each speech trial one of the 25 different Japanese narration sequences was randomly chosen. Overall, 25 trials were collected for each condition. The experiment started with 12 practice trials, but only the first two were declared so to the participant. This was to ensure that subjects invest effort to find a suitable strategy from the very beginning. The entire stimulus presentation and EEG triggering was done using Psyscope 1.2.5 (MacWhinney et al., 1997; see also http://psy.ck.sissa.it) running under Mac OS 9.

Data Acquisition and Pre-processing

EEG data were recorded from 64 channels (SynAmps, Neuroscan). Care was taken to make sure impedance did not exceed 5 kΩ. The data were sampled at 500 Hz and filtered online using a 0.1-200 Hz bandpass. During acquisition, the vertex electrode (Cz) served as reference, which was changed to average reference during offline-analysis. The multiple-source eye-correction method proposed by Berg and Scherg (1994) was used to correct artifacts caused by blinks. For ERP analysis, further epochs contaminated by artifacts were excluded by using a semi-automatic procedure offered in BESA (MEGIS, München). Wavelet and power-spectrum analyses were performed in Matlab, partly using custom-made functions as well as others taken from the eeglab toolbox (Delorme and Makeig, 2004).

ERP Analysis

After applying a low-pass filter (20 Hz) to the data, epochs of 1500 ms length around each visually presented digit were extracted from the data (300 ms pre-stimulus). In order to choose time-windows for exploration, the global power of each condition was calculated using all 64 channels for each individual. One thousand bootstrap replicates of paired t-tests were calculated for each sampling point between the start of stimulus presentation until 800 ms after the stimulus terminated. The latter was conducted for the following pairs of experimental conditions: Japanese versus silence as well as noise versus silence. A significant (95% confidence interval) sampling point for the Japanese-versus-silence comparison was only accepted if the comparison between noise and silence was not significant. Generally, the bootstrap statistic returns a probability distribution for the value (T) of a statistic (T), which can be used to calculate the confidence limits for the population parameter θ. The value T is obtained by sampling data of individuals with replacement. In this case an ordinary non-parametric bootstrap was applied (Davison and Hinkley 1997), thus using the empirical distribution to calculate all probability distribution parameters.

Wavelet Analysis

The artifact-corrected data were also submitted to a wavelet analysis (using eeglab; Delorme and Makeig, 2004; see also http://sccn.ucsd.edu/eeglab/). This function calculates the event-related spectral perturbations (ERSP), i.e. changes in spectral power (in dB) over time in a specified frequency band relative to the time-locking event (here: digit presentation) for each trial. These estimates are then averaged across trials. At the end, the mean baseline log power spectrum is subtracted from each spectral estimate yielding the baseline-normalized ERSP.

Time-frequency decompositions were performed by convoluting the time-series at a channel with hanning-tapered sinusoidal wavelets. The lowest frequency in our study was 0.06 Hz and the highest 89.38 Hz with a frequency resolution of 2.03 Hz.

Power Analysis

In order to analyse the spectral power in the retention phase, 8 s of movement-corrected raw data (first and last seconds omitted) were entered into a mean FFT analysis: Segments of continuous data (256 points) were tapered by a hanning window, followed by FFT calculation. The analysis window was moved by half a window length and the procedure was subsequently repeated. The mean FFT was then obtained by averaging the complex numbers of all time windows. Since we were particularly interested in paralleling our neurophysiological data with behavioural data, the spectral power of both sound conditions was related to the spectral power of the silence condition by

\[
\text{dB}_{i,e} = 20 \times \log_{10} \left( \frac{\text{SP}_{\text{stimulus}(i,e)}}{\text{SP}_{\text{baseline}(e)}} \right)
\]

where \(\text{SP}\) is the spectral power for a specific frequency (i) and electrode (e), thus yielding a sound-effect measure in dB.

Statistics

For the ERP and Wavelet Analyses, the significance of condition effects was determined using linear mixed-effects (LME) models (Pinheiro and Bates, 2000), which are implemented as a package running under R (R Development Core Team, 2004). This statistical method allows the modelling of data by incorporating fixed (associated with the population) and random (associated with the individual experimental unit) effects. The sound condition was specified as fixed effect, whereas the intercept and subject were specified as random effects.

Since there were no a-priori-hypotheses concerning the locations of presumed effects, the LME statistic was calculated for all 64 electrodes. Due to multiple testing, an α-adjustment was needed. Since a standard (Bonferroni) correction approach would have lead to an unacceptable loss of power (α = 0.0007), we decided to estimate an appropriate α from the data by means of a permutation test (Karniski et al., 1994). In this test, values for all three conditions for each subject at each electrode were resampled and the LME was calculated yielding 64 P-values. The single minimum value was stored from these P-values. This procedure was repeated 3000 times, thus leading to a distribution of 3000 minimal P-values. Subsequently, the bottom 5% quantile in the empirical distribution was chosen as α.

To test significant spectral power changes for the theta band compared to silence, dB values (cf. formula 1) for speech and noise were entered into a bootstrap statistic. In performing 3000 one-sample t-tests we tested whether the mean dB value significantly deviated from the suggested mean \(M = 0\) (95% confidence interval).

Results

Behavioural

The average performance registered for all 20 subjects is displayed in Figure 1. Strong performance deterioration under the speech condition for almost all serial positions (mean \(±\ SE = 42 ± 1.9\%\)) is shown, while the noise and silence conditions
almost yield identical results (32 ± 1.5 and 31 ± 1.5% respectively). The LME statistic shows significant sound condition \( F(2.38 = 9.92, P < 0.0003) \) and serial position effects \( F(1.477 = 298.51, P < 0.0001) \). The condition-by-position interaction failed to reach statistical significance \( F(2.477 = 0.85, P < 0.43) \), indicating that the effect evident in Figure 1 is merely due to an upward shift in the error curve for the speech condition.

**ERP**

Overall there appear to be three temporal clusters (with 4–5 sampling points) fulfilling the criteria of differentiating significantly between Japanese and silence but not between noise and silence: 96–116, 156–176 and 580–600 ms (see Fig. 2). Based on their topography, they will be termed P1, N1 and P6.

P1 is characterized by a bilateral posterior positivity and a central maximal negativation indicating at least two active sources in extrastriate areas. The difference map (not displayed here; available upon request) shows the positivation is more pronounced in the Japanese speaker condition. By contrast, N1 topography is dominated by a bilateral posterior pattern with negative potentials (see Fig. 3). Compared with P1, the centroids are located slightly towards the anterior, which implies that at least two sources are shifted in an anterior direction. The difference maps (lower panel, Fig. 3) show that the posterior negativation is reduced in the Japanese-speaker condition. Approximately 600 ms after stimulus onset, the ERP topography is marked by a centro-parietal postivitation (thus termed P6) and an anterior negativation. The comparison between speech and silence indicates that this pattern is more pronounced during speech.

#### Wavelet Analysis

In a first step four electrode clusters were formed for explorative purposes: left posterior, right posterior, left frontal and right frontal. Posterior clusters basically reflected the pattern of results gained with the ERP analysis. At frontal clusters, a considerable increase in gamma (20–80 Hz) was identifiable for speech and noise at ~300–400 ms, but it did not differentiate between these two conditions (not displayed here; available upon request). In the sub-20 Hz frequency range, an increase in power was observed for all clusters in the low frequency bands (<10 Hz) ~150–200 ms post-stimulus onset, corresponding to N1 latency. This is followed by a long-lasting decrease in power in a frequency band including the alpha and beta range (10–20 Hz). The most noticeable aspect distinguishing speech from the other sound conditions appears to be in the theta range (4–6 Hz). In a time window 400–800 ms after stimulus onset, speech shows less theta activity. Thus the mean theta activity in this time window was calculated for each individual and electrode, and then entered into the same LME and permutation statistic as described above. One right prefrontal electrode (Fp2) showed a lower \( P \) value than the one determined by the permutation test. The temporal evolution of theta activity for this electrode is displayed in the upper panel of Figure 5. An early theta reduction (pre-300 ms) can be seen for speech in comparison to the other two conditions, which is attributable to reduced N1. The reduction at 400–800 ms is also clearly evident from the figure. No corresponding effect could be seen in the ERP, and the display of inter-trial coherence (Fig. 5, lower panel) shows that phase-locking was rather low. We thus assume the late effect for theta seen in the wavelet analysis does reflect not strictly phase-locked (induced) neuronal activity.

**Spectral Power Analysis**

As described above, power changes (in dB) in the retention phase for the sound conditions were analyzed in comparison
Figure 3. The top panel displays the topography 156–176 ms post-stimulus onset for the three conditions. They are characterized by a marked posterior negativity. The difference maps for the sound versus silence comparison show an enhanced posterior negativity for the Japanese versus Silence condition only.

Figure 4. The headplots mark electrodes yielding significant effects according to a $P$-value gained via a permutation test. One can observe a left-lateralized posterior and anterior cluster, which presumably is due to the polarity reversal caused by an occipitoparietal source. The point-chart displays separately the mean (+SE) voltage for these electrode clusters.
with the silence condition. Values for the same frequency band, showing speech specific effects in the wavelet analysis (i.e. 4–6 Hz) at each electrode, were then submitted to the bootstrap test (see methods) in order to investigate whether they significantly deviate from zero. The upper panel of Figure 6 shades the area of electrodes significantly deviating from zero. This includes a cluster of right prefrontal-frontotemporal electrodes for the speech condition (left side of Fig. 6, top panel), which is not the case for the noise condition. The lower panel shows the mean power change for this electrode cluster. On average, both irrelevant sound conditions lead to a decrease in theta power, its change being considerably larger for the speech condition.

Discussion

The main intention of this EEG-study was to compare the effects of irrelevant speech and continuous noise on the neuronal processing of visually presented items in a serial recall task. We assumed that differences on a neuronal level which match the pattern of effects observed on a behavioural level could be of relevance for ISE generation. Our approach included ERPs and wavelet analyses. Additionally, we also investigated spectral power changes in the retention phase.

On a behavioural level, our data replicate the frequently reported significant detrimental effect irrelevant speech has on the serial recall of visual items (e.g. Buchner et al., 1996; Salamé and Baddeley, 1986, 1987), whereas noise leaves memory performance virtually unimpaired (e.g. Salamé and Baddeley, 1987; Jones et al., 1990; Ellermeier and Zimmer, 1997). Our main corresponding neuroelectrical findings are on the one hand a reduced visual N1 (~156–176 ms), and on the other hand reduced frontal theta activity following presentation of the to-be-remembered item (4–6 Hz; ~400–800 ms post-stimulus onset) as well as during the silent retention phase. Let us first consider the latter.

Theta oscillations are perhaps the most heavily investigated frequency band in neurophysiological working memory literature. Their general positive relationship with learning has been established in several animal studies (see Kahana et al., 2001). It could be shown, for example, that the possibility of inducing long-term potentiation is tightly linked to the hippocampal theta rhythm (e.g. McCartney et al., 2004). An increase in theta power during cognitive task execution has been linked to the encoding of new information into episodic memory by Klimesch (1999). According to this view, scalp recorded theta oscillations during memory performance reflect activity derived from cortico-hippocampal feedback loops, especially between hippocampus and limbic areas (Klimesch, 1996; Gevins et al., 1997, 1999; Sederberg et al., 2003). One notion holds that theta reflects a timing mechanism (Lisman and Idiart, 1995; Jensen et al., 1996): the assumption is that memory items are activated in gamma frequency subcycles of a theta oscillation. This idea is partly supported, on the one hand, by observations of increased theta activity during stimulus presentation in a Sternberg task, which persists during the retention phase (Raghavachari et al., 2001). On the other hand, theta is positively associated with working memory load (Jensen and Tesche, 2002). Also, a recent animal study (Lee et al., 2005) impressively demonstrated that neurons’ single unit activity involved in a delayed matching-to-sample task was phase-locked to certain periods of a theta cycle as measured by the local field potential. Generally, most studies are consistent and illustrate that oscillations in this frequency band are positively related to short-term memory performance.
show that during a proactive interference task, activity in right extrastriate occipito-parietal areas (DiRusso et al., 2001) was reported in several animal experiments using a delayed matching-to-sample paradigm (Fuster, 1973). Activity in this area alters with attentional demands, and it is assumed that this area is involved in the control of focussing attention (DeSouza and Everling, 2004). In a neuroimaging study, Uhl et al. (1994) could show that during a proactive interference task, activity in right prefrontal areas increases for high, but not for low interference conditions. These contributions also suggest the theta effect described above could be interpreted in favour of the assumption that irrelevant speech reduces directed attention to the to-be-remembered items.

Suggesting that cognitive mechanisms related to attention do in fact contribute to the ISE is in line with recently reported behavioural and neurophysiological data as mentioned in the introduction. However, attributing the ISE to attentional distraction is not a new idea (Broadbent, 1979; Cowan, 1995). Yet taking the major contributions on theoretically explaining the ISE into consideration, attentional aspects are at present only incorporated in the feature model’s explanation of the ISE (Neath, 2000). The phonological loop and the O-OER model—which have been extensively elaborated regarding the detrimental effect of irrelevant sounds on serial recall, too explicitly assume that attentional resources do not play any role in explaining the ISE. At this point it must be acknowledged that Norris et al. (2004) recently proposed implementing the ISE into their primacy model (Page and Norris, 1998). They also assume that irrelevant speech reduces, amongst other things, the amount of overall cognitive and attentional resources available (see Page and Norris, 2003; Norris et al., 2004). However, the primacy model currently lacks elaborate evaluation and testing regarding its explanation of the ISE, and therefore it still remains unclear whether it is suitable to explain and model this effect.

Admittedly, the role ascribed to attention in the feature model is not to degrade feature information, particularly not modality-dependent feature information as described in the introduction. This does not conform to our N1 effect which shows irrelevant speech can impair the basic sensory representation of the visual stimulus. Yet, based on the current data, we cannot claim with certainty whether loss of sensory feature representations due to irrelevant sound is behaviourally relevant, and whether it should be incorporated into cognitive models for the ISE. A promising approach to enlighten this aspect would be to find experimental manipulations which modulate the ‘ISE-sensitive’ EEG-variables found in this study along with performance. One possibility would be, for example, to manipulate irrelevant background sound by adding varying levels of noise to irrelevant speech, which systematically alters ISE magnitude (Ellermeier and Hellbrück, 1998).

Another question still left unanswered by this study is whether irrelevant speech has any particular effects on serial verbal recall which disturbing non-speech sound does not have, as proposed by feature model (Neath, 2000; Neath and Surprenant, 2001). A logical follow-up experiment would be to include a changing-state non-speech condition instead of the continuous noise condition. In doing so it would be possible to clarify whether irrelevant speech affects different aspects of ERP or oscillatory activity compared to changing-state non-speech.

Note here that a general problem arises when relating cognitive ISE models with neuroscientific approaches: they hardly overlap. Abstract concepts such as the previously mentioned ‘modality-dependent information’ are difficult to specify on a neurophysiological level and we can not be sure whether this term is understood in behavioural models the same way as in neurophysiological studies. However, as the present study shows, integrating the two approaches may be useful since evidence for the detrimental effect of irrelevant

Our observation of reduced theta activity in the speech condition (along with worse performance) is therefore in line with the findings outlined above.

In addition, speech-specific effects were already found at an early stage in the ERP for the visually presented items, namely the N1. The N1 is related to the processing of stimulus features in extrastriate occipito-parietal areas (DiRusso et al., 2001). A large amount of data show that this component is strongly modulated by directing attention to stimulus features or locations (e.g. Hillyard and Anllo-Vento, 1998; Hillyard et al., 1998): focussing attention on the stimulus leads to enhancements of the N1, thought to be associated with amplified stimulus processing, while distraction leads to reductions. In the light of these studies, our EEG data lead us to the assumption that speech automatically captures attentional resources or impairs focussing attention on the visually presented item. A very elegant psychophysical demonstration of how selective attention can alter feature processing has been recently undertaken by Carrasco et al. (2004).

Although inferences on underlying brain structures have to be made with great caution, it is tempting to assume that our theta results reflect reduced activity in right prefrontal cortical areas. This is corroborated by Gisselgård et al. (2003), who observed reduced metabolic activity in right frontal areas in an ISE experiment. Enhanced and enduring prefrontal activity was reported in several animal experiments using a delayed matching-to-sample paradigm (Fuster, 1973). Activity in this area alters with attentional demands, and it is assumed that this area is involved in the control of focussing attention (DeSouza and Everling, 2004). In a neuroimaging study, Uhl et al. (1994) could show that during a proactive interference task, activity in right prefrontal areas is increased for high, but not for low interference conditions. These contributions also suggest the theta effect described above could be interpreted in favour of the assumption that irrelevant speech reduces directed attention to the to-be-remembered items.

Suggesting that cognitive mechanisms related to attention do in fact contribute to the ISE is in line with recently reported behavioural and neurophysiological data as mentioned in the introduction. However, attributing the ISE to attentional distraction is not a new idea (Broadbent, 1979; Cowan, 1995). Yet taking the major contributions on theoretically explaining the ISE into consideration, attentional aspects are at present only incorporated in the feature model’s explanation of the ISE (Neath, 2000). The phonological loop and the O-OER model—which have been extensively elaborated regarding the detrimental effect of irrelevant sounds on serial recall, too explicitly assume that attentional resources do not play any role in explaining the ISE. At this point it must be acknowledged that Norris et al. (2004) recently proposed implementing the ISE into their primacy model (Page and Norris, 1998). They also assume that irrelevant speech reduces, amongst other things, the amount of overall cognitive and attentional resources available (see Page and Norris, 2003; Norris et al., 2004). However, the primacy model currently lacks elaborate evaluation and testing regarding its explanation of the ISE, and therefore it still remains unclear whether it is suitable to explain and model this effect.

Admittedly, the role ascribed to attention in the feature model is not to degrade feature information, particularly not modality-dependent feature information as described in the introduction. This does not conform to our N1 effect which shows irrelevant speech can impair the basic sensory representation of the visual stimulus. Yet, based on the current data, we cannot claim with certainty whether loss of sensory feature representations due to irrelevant sound is behaviourally relevant, and whether it should be incorporated into cognitive models for the ISE. A promising approach to enlighten this aspect would be to find experimental manipulations which modulate the ‘ISE-sensitive’ EEG-variables found in this study along with performance. One possibility would be, for example, to manipulate irrelevant background sound by adding varying levels of noise to irrelevant speech, which systematically alters ISE magnitude (Ellermeier and Hellbrück, 1998).

Another question still left unanswered by this study is whether irrelevant speech has any particular effects on serial verbal recall which disturbing non-speech sound does not have, as proposed by feature model (Neath, 2000; Neath and Surprenant, 2001). A logical follow-up experiment would be to include a changing-state non-speech condition instead of the continuous noise condition. In doing so it would be possible to clarify whether irrelevant speech affects different aspects of ERP or oscillatory activity compared to changing-state non-speech.

Note here that a general problem arises when relating cognitive ISE models with neuroscientific approaches: they hardly overlap. Abstract concepts such as the previously mentioned ‘modality-dependent information’ are difficult to specify on a neurophysiological level and we can not be sure whether this term is understood in behavioural models the same way as in neurophysiological studies. However, as the present study shows, integrating the two approaches may be useful since evidence for the detrimental effect of irrelevant
background speech on the feature representation of the to-be-remembered items was produced. This effect is most likely due to a reduction of attention to the visually presented memory items occurring during both item presentation and rehearsal.

Notes
The authors would like to express their appreciation to Anke Trez and Thomas Hartmann for their support during data collection. We would also like to thank two anonymous reviewers and Cara Kahl for their many helpful suggestions on an earlier draft of this article.

Address correspondence to Nathan Weisz, University of Konstanz, Department of Psychology, PO Box 25, D-78457 Konstanz, Germany. Email: Nathan.Weisz@uni-konstanz.de.

References