Word frequency of irrelevant speech distractors affects serial recall

AXEL BUCHNER
Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany

and

EDGAR ERDFELDER
Universität Mannheim, Mannheim, Germany

In this study, participants memorized frequent or rare target words in silence or while ignoring frequent or rare distractor words. Distractor words impaired recall performance, but low-frequency distractor words caused more impairment than did high-frequency distractor words. We demonstrate how to solve the identifiability problem for Schweickert's (1993) multinomial processing tree model of immediate recall, and then use this model to show that irrelevant speech affected both the probability with which intact target word representations were available for serial recall and the probability of successful reconstruction of item identities based on degraded short-term memory traces. However, the type of irrelevant speech—low- versus high-frequency words—selectively affected the probability of intact target word representations. These results are consistent with an explanation of the irrelevant speech effect within the framework proposed by Cowan (1995), and they pose problems for other explanations of the irrelevant speech effect. The analyses also confirm the validity of Schweickert's process model.

Task-irrelevant auditory stimuli impair serial recall of visually presented words. A substantial body of knowledge has been accumulated about the variables that do and do not affect this "irrelevant sound effect" (Baddeley, 2000; Jones & Tremblay, 2000; Neath, 2000). For instance, the size of the irrelevant sound effect seems to be independent of the phonological and semantic similarity between the to-be-remembered material and the irrelevant material, and for quite some time it seemed to be independent of whether irrelevant spoken material is meaningful or not (Buchner, Irmen, & Erdfelder, 1996; Colle & Welsh, 1976; LeCompte & Shaibe, 1997; Salamé & Baddeley, 1982). In contrast, the irrelevant material's acoustic properties, particularly the changing-state property, play a substantial role (Jones, 1993; Jones, Beaman, & Macken, 1996). We report evidence of another variable that affects immediate recall performance—that is, the frequency of auditory distractor words in the language.

Given that immediate serial recall must rely heavily on working memory processes, theories of working memory play an important role in explaining the irrelevant sound effect. These theories fall into one of two categories—theories that explicitly specify a role for atten-

tion in the irrelevant sound effect and theories that do not but instead assume that irrelevant sounds have automatic access to the representational structure that is also used for the primary task of maintaining a suitable representation of the to-be-recalled words (cf. Elliott, 2002). An exemplar of the latter category is Baddeley's modular working memory model (Baddeley, 1986, 1996; Baddeley & Logie, 1999). Within this model, it is assumed that the preferred strategy for the immediate serial recall, at least of short word lists, is to convert visually presented words into a phonological representational format so that the words can be maintained in the limited-capacity phonological loop module of working memory. Irrelevant auditory speech is said to gain automatic access to this store where it interferes with the target representations, which results in the impairment of serial recall (Baddeley, 1986; Salamé & Baddeley, 1982, 1989). The impairment is assumed to occur in a structure in which the representational format is exclusively phonological. Therefore, nonacoustic distractor properties must not affect serial recall performance. The above mentioned irrelevance of the semantic content of the irrelevant distractor information for the size of the irrelevant speech effect seemed consistent with this prediction, although more recent research showed that semantic content may adversely affect recall performance (Buchner, Rothermund, Wentura, & Mehl, 2004; Neely & LeCompte, 1999). Importantly for the present context, word frequency as a nonacoustic distractor feature also should not affect serial recall performance.

The same prediction can be derived from the object-oriented episodic record model (Jones, 1993; Jones & Macken, 1993). Within this model, the crucial process in immediate serial recall is that of seriation of the to-be-recalled objects. These objects are temporarily assembled on a structure referred to as a *blackboard*, where they are linked by a series of production rules. Objects from the visual domain—the to-be-recalled target words in the present task—and the links interconnecting them are constructed by means of articulation, but once on the blackboard, objects are represented in a code in which they are undifferentiated in terms of their modality of origin. Auditory objects—the distractor words in the present case—result from preattentive segmentation processes. Irrelevant auditory material disrupts serial recall of visual objects because a set of competing links is established automatically among auditory objects, and the links interfere with those for the visual objects. In other words, it is the integrity of the links, not the integrity of the item representations per se, that limits serial recall performance; if a link to an item is lost, the item is no longer available for recall. Perhaps the greatest strength of the object-oriented episodic record model is that it correctly predicts the changing-state effect—that is, a more disruptive effect of irrelevant sound that is characterized by a greater degree of segmentation (Jones & Tremblay, 2000). Importantly for the present research, the frequency of words in a language cannot determine the number of changes inherent in the auditory signal, which is why word frequency of irrelevant distractors should not affect serial recall performance.

Within Cowan's (1995, 1999) attention-and-memory framework, working memory is functionally defined by the set of cognitive processes that are needed to retain information stored in long-term memory in a highly accessible state. The focus of attention represents the most highly activated subset of objects in working memory. In serial recall tasks such as the one considered here, the rehearsed target items represent the focus of attention. Irrelevant auditory distractors may interfere in two ways with those targets. First and foremost, distractors may automatically attract attention, thereby recruiting processing resources away from the currently attended targets. A reduction of attention available for rehearsing the target items reduces their activation levels and, hence, the probability of successful recall of the targets. Second, given that interference occurs between representations in memory that are similar, acoustic distractors may interfere with the representations of target items that were recoded phonologically. Within this framework, the type of auditory distractor could affect serial recall performance if one type of distractor were to recruit more attention, leaving fewer resources for the rehearsal processes that constitute the present focus of attention. Thus, provided that the processing of very frequent words is less resource demanding than the processing of extremely rare words, the latter could impair recall even more than the former.

The feature model (Nairne, 1990; Neath, 2000) assumes that the features of the irrelevant sounds overwrite a certain number of elements of the feature vectors of the targets, which degrades the target representations in working memory. With fewer intact features, there is a reduction in the probability of successfully matching a degraded target item representation to the representations in long-term memory. Recall depends on a successful match, which is why irrelevant speech should reduce the probability of successful recall from working memory. The feature model also includes an attentional parameter that can be changed to reflect the amounts of processing resources available for the memorization task. If the processing of very rare words attracted more of a limited attentional resource than the processing of very frequent words, ignoring the former type of distractors should leave relatively less attention for the memorization task, resulting in serial recall performance that is even worse for rare than it is for frequent words. In addition, as Neath (2000) has pointed out, the feature model allows deriving the prediction that the factors of irrelevant speech and serial position do not interact. This is necessarily so because within the feature model, the overwriting of target features is assumed to occur irrespective of the targets' serial positions.

To this end, we derived predictions from competing theories of working memory for the effects of the frequency of irrelevant words on raw serial recall performance. In a next step, we want to derive predictions more directly for processes that can be assumed to underlie serial recall. A multinomial model specifying these processes has been proposed by Schweickert (1993) and was evaluated by Hulme et al. (1997). This model should not be seen as a separate theory of working memory, competing with the other theories. Rather, it should be viewed as a "measurement model" (e.g., Batchelder & Riefer, 1999; Riefer & Batchelder, 1988)—that is, as a tool that provides for measuring the cognitive processes that are relevant for evaluating each of the working memory theories discussed so far.

Figure 1 illustrates Schweikert's (1993) model. It assumes that, at the time of recall, an intact representation of a target item may exist so that it can be recalled without hesitation and correctly. This occurs with probability i_s, where s represents the serial position of the item. By assuming a separate i_s parameter for each serial position s, the model allows the probability of an existing intact memory trace to differ among serial positions. With probability $1-i_s$, there is no intact item representation. However, a degraded memory trace may still exist on the basis of which the item identity can be "reconstructed," with probability r_j, where j represents an experimental situation with a certain overall probability of successful reconstruction that is the same for items at all serial positions. It follows that parameter r_j should be affected by variables that influence the degree of difficulty of reconstructing targets from degraded short-term memory traces, which is a function of both the degree of degra-
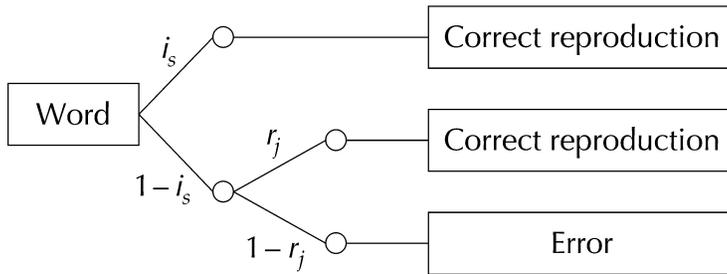
Figure 1. Graphic illustration of the multinomial model of serial recall processes suggested by Schweickert (1993).

dation of the target's short-term memory trace and the accessibility of the long-term memory representation against which it is matched. For instance, frequent words should have more accessible long-term memory representations than do rare words. They should therefore be reconstructed more readily than rare words, which is why parameter r_j is predicted to be larger for frequent than for rare words. In contrast, parameter r_j should *not* vary as a function of variables that influence the number of intact item representations in short-term memory. However, such variables are predicted to affect the i_s parameters that represent the probability of the existence of an intact item representation. For instance, assuming a limited-capacity auditory short-term memory, it can be predicted that more intact representations of short than of long words can be maintained by rehearsal. Thus, parameters i_s, but not parameters r_j, should vary as a function of variables such as word length.

The predictions of the model can be illustrated using the data from Experiment 3 of Hulme et al. (1997), which are reproduced in Table A1 of Appendix A. Participants memorized series of seven words that were sampled without replacement from one of four eight-word sets: short-length high-frequency words; short-length low-

frequency words; medium-length high-frequency words; and medium-length low-frequency words. In essence, participants recalled a larger number of short- than of medium-length words, they recalled a larger number of frequent than of rare words, and these two effects were more pronounced at intermediate serial positions.

Although Hulme et al. (1997) referred to Schweickert's (1993) model when analyzing their results, they did not fit the model to the complete data sets of their experiments, and they did not test specific hypotheses using appropriate restrictions imposed on model parameters. We report such results, using the data from their Experiment 3, for two reasons. First, these results confirm the model's validity. Second, the design of the experiment reported below is similar to Hulme et al.'s Experiment 3, so their results are particularly relevant for the present purposes.

Obtaining unique parameter estimates, correct goodness-of-fit statistics, and appropriate degrees of freedom for Schweickert's (1993) multinomial model is not straightforward, however. Because two model parameters are used for fitting a single frequency of correct recall, one easily runs into problems of parameter identifiability. In Appendix B, we discuss ways of coping with this problem and

Table 1
Global Model Fit and Hypothesis Tests of the Multinomial Model of Serial Recall
Processes Proposed by Schweickert (1993) When Applied to the
Data From Experiment 3 of Hulme et al. (1997)

Hypothesis	Goodness-of-Fit Statistic‡	p value
Global model fit (including two identifiability constraints)	$G^2(12) = 11.80$	$p = .46$
Hypothesis tests	$\Delta G^2(6) = 30.00$	$p < .0001$
$i_{1,\text{ short length}} = i_{1,\text{ medium length}}$*		
$i_{2,\text{ short length}} = i_{2,\text{ medium length}}$		
$i_{3,\text{ short length}} = i_{4,\text{ medium length}}$		
$i_{4,\text{ short length}} = i_{4,\text{ medium length}}$		
$i_{5,\text{ short length}} = i_{5,\text{ medium length}}$		
$i_{6,\text{ short length}} = i_{6,\text{ medium length}}$		
$i_{7,\text{ short length}} = i_{7,\text{ medium length}}$		
$r_{\text{short length, high-frequency}} = r_{\text{short length, low frequency}}$	$\Delta G^2(2) = 125.38$	$p < .0001$
$r_{\text{medium length, high frequency}} = r_{\text{medium length, low frequency}}$		
$r_{\text{short length, high frequency}} = r_{\text{medium length, high frequency}}$*	$\Delta G^2(1) = 0.02$	$p = .89$
$r_{\text{short length, low frequency}} = r_{\text{medium length, low frequency}}$		

*These parameter restrictions served as identifiability constraints (see Appendix B for details).   ‡The likelihood-ratio goodness-of-fit statistic G^2 is approximately χ^2 distributed.

suggest a method that works well for most applications. This method was used for all applications of Schweickert's (1993) model in the present article.

Table 1 displays the results of the global model fit as well as of the tests of specific substantive hypotheses for the data of Hulme et al. (1997, Experiment 3).[1] Obviously, the overall G^2 goodness-of-fit statistic for the general model containing 14 i_s parameters (one for each serial position and each word length) and 4 r_j parameters (one for each target word frequency and each word length) is excellent. The specific prediction that word length affects the probability of the existence of intact item representations and hence, in model terms, the i_s parameters, is tested by imposing the additional restrictions that, for each serial position s, parameter i_s is identical for short- and medium-length words. These additional restrictions generate six degrees of freedom (one for each of the seven serial positions minus one because one restriction already served as an identifiability constraint for the general model; see Appendix B), so that the restricted model has 18 degrees of freedom. The appropriateness of the additional set of restrictions is tested by subtracting the χ^2-distributed G^2 goodness-of-fit statistic for the global, unrestricted model, which has 12 degrees of freedom, from the χ^2-distributed G^2 goodness-of-fit statistic for the restricted model, which has 18 degrees of freedom. The resulting ΔG^2 difference test statistic is also χ^2-distributed with $18-12 = 6$ degrees of freedom.[2] As predicted, the hypothesis clearly must be rejected so that we may conclude that word length affects the probability of the existence of intact item representations. Word frequency, in contrast, should affect the reconstruction parameter r_j. Consistent with this prediction, the hypothesis that parameter r_j is identical for high- and low-frequency words must also be rejected. Finally, there is no reason for word length to affect the probability of a successful reconstruction process. The excellent fit of the model with the restriction that parameter r_j is the same for short- and medium-length words is clearly consistent with this prediction. This is parallel to conclusions arrived at by Schweickert, Chen, and Poirier (1999) on the basis of descriptive comparisons of the values of parameter estimates (although, as we show in Appendix B, absolute values of parameter estimates, cannot be interpreted as probability estimates and only ratios of differences between parameter estimates are uniquely determined by the data). In sum, then, we may conclude that Schweickert's multinomial model is valid and successfully explains serial recall data.

Our intention was to use this established model for the purpose of testing predictions derived from the models of the irrelevant speech effect discussed above. Basically, while participants memorized high- and low-frequency target words as in Hulme et al. (1997), no distractors (silent control condition), high-frequency distractor words, or low-frequency distractor words had to be ignored.

Our first goal was to replicate the basic findings of Hulme et al. (1997, Experiment 3), that the model fits the serial recall data (which implies that serial position selectively affects the i_s parameters, but not the r_j parameters, which are assumed to be identical across all serial positions) and that target word frequency affects the r_j parameters. Next, hypotheses about the effects of irrelevant speech could be tested. All working memory models considered in the introduction predict that, when compared with the silent control condition, irrelevant speech reduces the probability with which intact target representations are available for immediate recall. Thus, the i_s parameters should differ among the silent control and the two irrelevant speech conditions.

The models differ with respect to the predicted effect of distractor word frequency on the i_s parameters: Models that allow for attentional resources to play a role in the irrelevant sound effect (Cowan, 1995, 1999; Nairne, 1990; Neath, 2000) can predict smaller i_s parameters in the presence of low- rather than high-frequency distractor words because low-frequency distractor words require more processing resources that could otherwise have been used for keeping the memory representations of the target words active and intact. In contrast, the modular working memory model (Baddeley, 1986, 1996; Baddeley & Logie, 1999) as well as the object-oriented episodic record model (Jones, 1993; Jones & Macken, 1993) cannot account for effects of nonacoustic distractor features and thus predict a null effect of distractor word frequency on the i_s parameters.

The models also differ in their predictions for the r_j parameters. According to the object-oriented episodic record model, to-be-recalled items are linked, and auditory distractors interfere by establishing competing links (Jones, 1993; Jones & Macken, 1993). Without a link pointing to a target item, there is no mechanism by which it can be recalled. In other words, a target representation can only be accessible (when it is pointed to by a link) or not, which is why irrelevant speech according to this model should affect the i_s parameters as explicated above, but irrelevant speech should not affect the r_j parameters, which should be identical for the silent control and the two distractor conditions.

In contrast, the feature model (Nairne, 1990; Neath, 2000) predicts that the r_j parameters differ between the silent control and the two irrelevant speech conditions. This is so because the features of the irrelevant sounds are assumed to overwrite a certain number of elements of the feature vectors of the target representations, which cause these representations to become more degraded, and, hence, less likely to be amenable to successful reconstruction than target representations in the silent control condition.

The same prediction may be derived for the modular working memory model (Baddeley, 1986, 1996; Baddeley & Logie, 1999) and within the framework of Cowan (1995, 1999), although this deduction is somewhat less

stringent. In essence, interference between targets and distractors may occur to the degree to which these two item types share a representational format. If the term *interference* is taken to imply that target representations may become degraded by the presence of distractors, these two models' predictions for the r_j parameters are parallel to the feature model's predictions.

However, it seems that the modular working memory model would also be compatible with a different pattern of results. One could assume that, when auditory distractors enter the limited-capacity phonological store, target items are lost completely because of the limited number of items that can be held simultaneously in that store. According to this interpretation of the model, auditory distractors should only affect the i_s parameters as described above, but not the r_j parameters.

METHOD

Participants

The participants were 54 students (35 women) who were paid for their participation. Their age ranged from 19 to 30 years ($M = 23$). All participants were tested individually.

Materials

The participants wore headphones that were plugged directly into an Apple iMac computer, which controlled the experiment. Sounds were produced at a level of 72 ± 5 dB(A).

Word frequency was determined using the German language corpus available in the CELEX database (Centre for Lexical Information, 1991). The selected words, their frequencies in the corpus, and independent concreteness ratings of 127 students from the same population as the participants are reported in Table A3 of Appendix A.

The entire set of 32 six-letter, two-syllable words were divided into two sets of high- and low-frequency words. The four sets were matched, as well as possible, in average concreteness and in the frequencies of different types of vowels, while maximizing the difference in frequency between high- and low-frequency words. The two sets of high-frequency words were also matched, as well as possible, in word frequency, as were the low-frequency word sets.

For one randomly determined half of the participants, the visually presented target words were from Sets 1 (high frequency) and 3 (low frequency), and the auditory distractors were from Sets 2 (high frequency) and 4 (low frequency). For the other half of the participants, the assignment to the target and distractor categories was reversed.

The auditory distractors were spoken by a female voice and were digitally recorded at 44.1 kHz using 16-bit encoding. Each word was edited to last 700 msec and was normalized so as to minimize amplitude differences among the words.

Procedure

The experiment began with three practice sequences during which seven visually presented two-syllable training words had to be remembered for immediate verbal serial recall. Each word was presented for 700 msec. There was no pause between words. After a retention interval of 1 sec, the words had to be recalled in the order in which they had been presented. The participants responded with the German analogue to the word "blank" for each word they could not recall. The participants' responses were recorded by a tape recorder for later evaluation.

The 48 experimental sequences were parallel to the training sequences but consisted of seven visually presented words that were randomly drawn without replacement from the eight words of one of the word sets reported in Table A3 of Appendix A. These target words had to be remembered for immediate serial recall. Twenty-four of the sequences were composed of high-frequency target words, and 24 were composed of low-frequency target words. Within the 24 sequences of each target type, no auditory distractors were presented on 8 randomly selected sequences, high-frequency distractor words were presented aurally on another set of 8 randomly selected sequences, and low-frequency distractor words were presented aurally on a final set of 8 sequences. On sequences with distractors, seven of the 700-msec auditory distractor words were selected randomly without replacement from the set of eight words of the relevant word set. Thus, one auditory distractor was presented parallel to one visual target with no period of silence between distractor presentations.

On average, the experiment lasted about 40 min, after which the participants were informed about its purpose.

Design

The independent variables were target type (high- vs. low-frequency words; within-subjects), distractor type (silence, high-frequency words, and low-frequency words; within-subjects), and serial position. The dependent variable was the participants' serial recall performance, that is, the number of visually presented words correctly recalled at the serial position at which they were presented.

The most relevant independent variable was that of the distractor type. Given a total sample size of $N = 54$, $\alpha = .05$, and an "average" population correlation between the levels of this repeated measures factor of $\rho = .50$ (estimated from pilot data), distractor effects of size $f^2 = 0.25$ (medium size effects in terms of the conventions suggested by Cohen, 1977) could be detected with a probability of $1 - \beta = .98$. For planned contrasts between only two distractor type conditions (e.g., between high- and low-frequency distractors) and otherwise identical parameters, the power was $1 - \beta = .95$.[3] The level of α was set to .05 for all analyses reported in the present article. Partial R^2s are reported as a measure of the size of an effect.

RESULTS

Figure 2 illustrates the three basic properties of the data. First, performance differed as a function of the serial position. Second, high-frequency targets were recalled better than low-frequency targets. Third, recall was best when no distractors were present, clearly worse with high-frequency distractor words, and even worse with low-frequency distractor words. A repeated measures multivariate analysis of variance[4] showed statistically significant main effects of serial position [$F(6,48) = 141.27$, $R_p^2 = .95$], of target type [$F(1,53) = 35.84$, $R_p^2 = .40$], and of distractor type [$F(2,52) = 55.08$, $R_p^2 = .68$]. The target type effect replicates the results of Hulme et al. (1997). Planned orthogonal contrasts on the distractor type variable show that the silent condition differed significantly from the two conditions in which irrelevant auditory distractors were present [$F(1,53) = 83.09$, $R_p^2 = .61$] and that the difference in performance between high- and low-frequency distractor conditions was also significant [$F(1,53) = 20.84$, $R_p^2 = .28$].

There was also an interaction between serial position and distractor type [$F(6,48) = 5.76$, $R_p^2 = .62$], reflecting the fact that the difference between the conditions defined by the distractor type variable was reduced on the sixth serial position and was virtually eliminated on
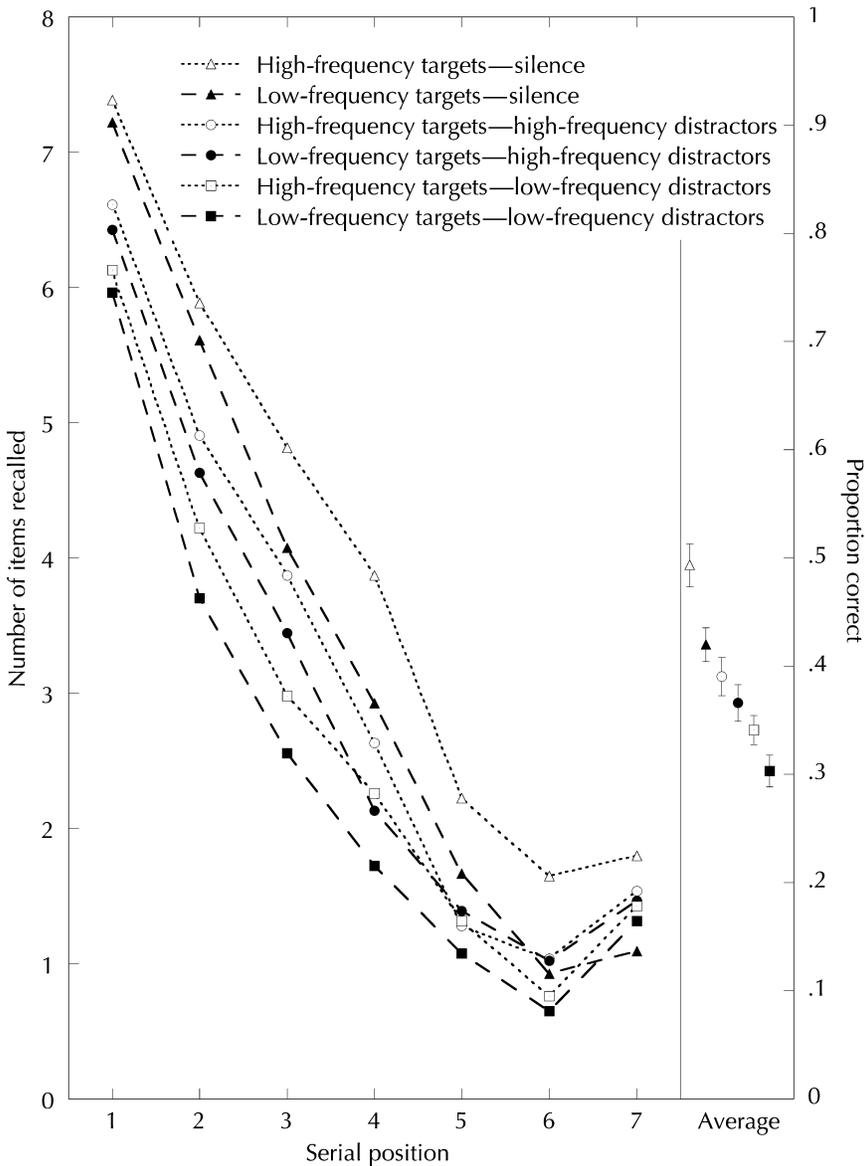
Figure 2. Cumulated number of words recalled correctly at each serial position (eight words at most) and proportion correct summed over serial positions.

the final serial position, but only for conditions with au-ditory distractors.

Neither the interaction between target and distractor type [$F(2,52) = 2.88$], nor the interaction between target type and serial position [$F(6,48) = 1.50$], nor the triple interaction [$F(12,42) = 0.63$] was significant.

The results of the analyses based on the multinomial model of immediate recall processes (Hulme et al., 1997; Schweickert, 1993), again using the method suggested in Appendix B, are summarized in Table 2 (model fit and hypothesis tests) and in Figure 3 (parameter estimates). First, the full model with 21 i_s parameters (one for each serial position and each distractor type condition) and 6

r_j parameters (one for each target word frequency and each distractor type condition) fits the data very well. Second, we replicated the finding of Hulme et al. (1997, Experiment 3) that parameter r_j varies as a function of the target word frequency: The model with the restric-tion that the r_j parameters do not differ between high- and low-frequency targets must be rejected. Third, all working memory models predict the i_s parameters to vary as a function of the distractor type conditions. In-deed, the model with the restriction that, for each serial position s, parameter i_s does not differ among the three irrelevant speech conditions must be rejected. These re-sults confirm the validity of Schweickert's process

Table 2
Global Model Fit and Hypothesis Tests of the
Multinomial Model of Serial Recall Processes Proposed by Schweickert (1993)
When Applied to the Data From the Present Experiment

Hypothesis	Goodness-of-Fit Statistic[‡]	p value
Global model fit (including 3 identifiability constraints)	$G^2(18) = 17.84$	$p = .47$
Hypothesis tests	$\Delta G^2(12) = 203.31$	$p < .0001$
$i_{1,\,\text{silence}} = i_{1,\,\text{high-frequency distractors}} = i_{1,\,\text{low-frequency distractors}}$ [*]		
$i_{2,\,\text{silence}} = i_{2,\,\text{high-frequency distractors}} = i_{2,\,\text{low-frequency distractors}}$		
$i_{3,\,\text{silence}} = i_{3,\,\text{high-frequency distractors}} = i_{3,\,\text{low-frequency distractors}}$		
$i_{4,\,\text{silence}} = i_{4,\,\text{high-frequency distractors}} = i_{4,\,\text{low-frequency distractors}}$		
$i_{5,\,\text{silence}} = i_{5,\,\text{high-frequency distractors}} = i_{5,\,\text{low-frequency distractors}}$		
$i_{6,\,\text{silence}} = i_{6,\,\text{high-frequency distractors}} = i_{6,\,\text{low-frequency distractors}}$		
$i_{7,\,\text{silence}} = i_{7,\,\text{high-frequency distractors}} = i_{7,\,\text{low-frequency distractors}}$		
$i_{1,\,\text{high-frequency distractors}} = i_{1,\,\text{low-frequency distractors}}$ [*]	$\Delta G^2(6) = 50.47$	$p < .0001$
$i_{2,\,\text{high-frequency distractors}} = i_{2,\,\text{low-frequency distractors}}$		
$i_{3,\,\text{high-frequency distractors}} = i_{3,\,\text{low-frequency distractors}}$		
$i_{4,\,\text{high-frequency distractors}} = i_{4,\,\text{low-frequency distractors}}$		
$i_{5,\,\text{high-frequency distractors}} = i_{5,\,\text{low-frequency distractors}}$		
$i_{6,\,\text{high-frequency distractors}} = i_{6,\,\text{low-frequency distractors}}$		
$i_{7,\,\text{high-frequency distractors}} = i_{7,\,\text{low-frequency distractors}}$		
$r_{\text{high-frequency targets, silence}} = r_{\text{low-frequency targets, silence}}$	$\Delta G^2(3) = 57.97$	$p < .0001$
$r_{\text{high-frequency targets, high-frequency distractors}} = r_{\text{low-frequency targets, high-frequency distractors}}$		
$r_{\text{high-frequency targets, low-frequency distractors}} = r_{\text{low-frequency targets, low-frequency distractors}}$		
$r_{\text{high-frequency targets, silence}} = r_{\text{high-frequency targets, high-frequency distractors}} = r_{\text{high-frequency targets, low-frequency distractors}}$ [*]	$\Delta G^2(2) = 21.04$	$p < .0001$
$r_{\text{low-frequency targets, silence}} = r_{\text{low-frequency targets, high-frequency distractors}} = r_{\text{low-frequency targets, low-frequency distractors}}$		
$r_{\text{high-frequency targets, high-frequency distractors}} = r_{\text{high-frequency targets, low-frequency distractors}}$ [*] $= r_{\text{low-frequency targets, high-frequency distractors}} = r_{\text{low-frequency targets, low-frequency distractors}}$	$\Delta G^2(1) = 0.85$	$p > .50$

[*]These parameter restrictions served as identifiability constraints (see Appendix B for details). [‡]The likelihood-ratio goodness-of-fit statistic G^2 is approximately χ^2 distributed.

model, allowing us to conclude that irrelevant speech affects the probability with which intact short-term memory traces are available for immediate recall. Next, we report the specific tests that help to differentiate among working memory models.

First, the i_s parameters were smaller for low-frequency distractors, as compared with high-frequency distractors, as is indicated by the significant ΔG^2 statistic obtained for the model imposing equality constraints on the i_s parameters for the high- and low-frequency distractors only. This result contradicts the predictions based on the modular working memory model (Baddeley, 1986, 1996; Baddeley & Logie, 1999) and the object-oriented episodic record model (Jones, 1993; Jones & Macken, 1993) but is consistent with the predictions derived from Cowan's (1995, 1999) framework, as well as the feature model (Nairne, 1990; Neath, 2000).

Second, the additional restriction that the r_j parameters do not differ as a function of the irrelevant speech condition also must be rejected. This is inconsistent with the object-oriented episodic record model (Jones, 1993; Jones & Macken, 1993) but is consistent with the other models considered here.

Third and finally, imposing equality constraints only on the r_j parameters for high- and low-frequency distractors does not produce a significant decrease in goodness-of-fit, so that the hypothesis of equal r_j parameters for

high- and low-frequency distractors can be maintained. Thus, we may conclude that irrelevant speech serves to degrade short-term memory representations, but it does so irrespective of the distractor word frequency.

DISCUSSION

The analyses based on the raw serial recall performance measure and the model-based analyses converge in that their results are clearly incompatible with working memory models that do not specify a role for attention. Specifically, the data are inconsistent with current formulations of the modular working memory model (Baddeley, 1986, 1996; Baddeley & Logie, 1999) and the object-oriented episodic record model (Jones, 1993; Jones & Macken, 1993). These models predict that nonacoustic distractor properties must not affect serial recall performance and that, in terms of Schweickert's (1993) process model of immediate recall, parameters i_s must not vary as a function of distractor word frequency. Both of these predictions were disconfirmed. The object-oriented episodic record model also predicts that a memory trace in short-term memory is either accessible (when a link points to it) or not (when an auditory distractor successfully competes for that link). In other words, traces cannot be degraded, which is why this model predicts that parameter r_j, in contrast to parameter i_s, should remain unaffected by the pres-
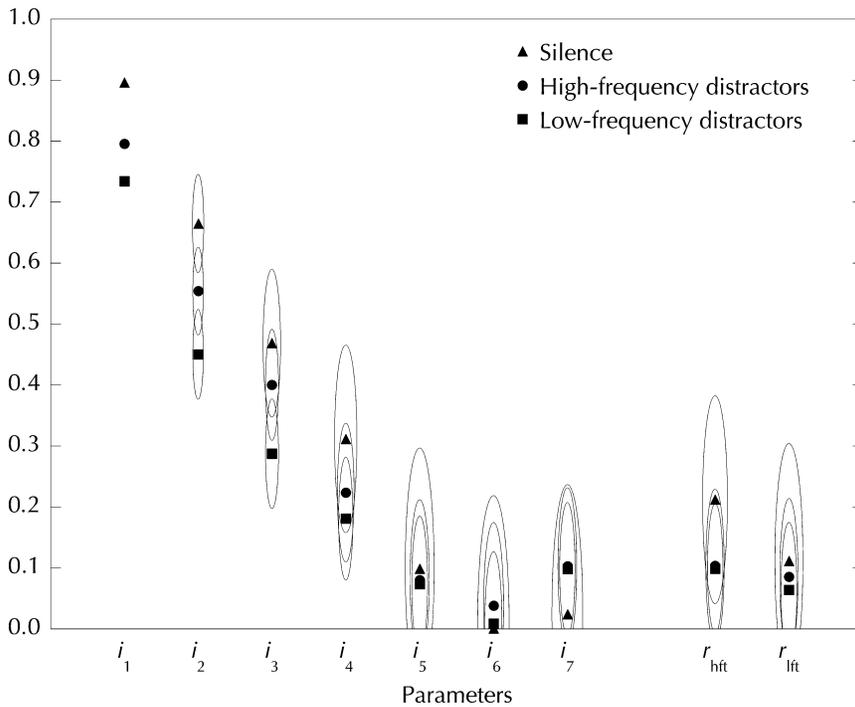
Figure 3. Parameter estimates for Schweickert's (1993) multinomial model specifying the processes underlying serial recall. Intact representations of a target item are denoted by the i_s parameters, where s represents the serial position of the item. The i_1 parameters for the three distractor type conditions were fixed at constant values so as to serve as identifiability constraints. Successful reconstructions from degraded memory traces are denoted by the r parameters, where hft and lft represent the reconstructions of high- and low-frequency targets, respectively. Error arcs represent .95 confidence intervals. Note that parameter estimates are only unique up to linear transformations of the type $i_s': = 1 - c(1 - i_s)$ and $r_j': = 1 - (1/c)(1 - r_j)$ (see Appendix B for details).

ence or absence of irrelevant auditory distractors. The present results contradict this expectation.

In contrast, the feature model (Nairne, 1990; Neath, 2000) is generally consistent with the results. However, the model predicts that the performance impairment due to irrelevant speech should occur uniformly across all serial positions. This prediction was disconfirmed by the finding that the size of the effect of irrelevant speech depended on the serial position. Note that this disconfirmation was possible only because the feature model makes a strong prediction, with respect to which the other models are simply silent. Also, although the interaction between irrelevant speech and serial position observed here and in other experiments from our lab (Buchner, Irmen, & Erdfelder, 1996; Buchner et al., 2004) is problematic for the feature model, the empirical situation pertaining to the interaction between irrelevant speech and serial position is inconsistent (cf. Neath, 2000) so that more evidence is needed before a firm judgment on this issue can be reached.

The framework proposed by Cowan (1995, 1999) is most compatible with the present results. It correctly predicted that rare auditory distractor words would impair serial recall performance even more than high-frequency

words because the probability i_s of an intact target word memory representation is further reduced if target words are learned in the presence of attention-attracting low-frequency distractors. Thus, we conclude that the results reported here seem best explained within this framework.

This conclusion converges with those reached earlier by others. For instance, Elliott (2002) found, in a developmental study, that the impairment in recall performance caused by irrelevant sounds decreased with increasing age. If we assume that attentional selection abilities get better as children grow older, this result supports the hypothesis that attention plays a role in the maintenance of information in working memory. Neely and LeCompte (1999) reported that the semantic similarity between targets and distractors plays a role for serial recall performance. This result cannot be accounted for by the modular working memory model and the object-oriented episodic record model, because both models imply that only acoustic distractor features affect serial recall performance. It is for the same reason that these models cannot account for the findings reported by Buchner et al. (2004), who reported that both positively and negatively valent distractor words caused more disruption of serial recall performance than did neutral dis-

tractors, and negative distractors caused more disruption than did positive distractors. Given that research using the Stroop task has shown that the attention-grabbing power of emotionally valent words is larger than that of neutral words, and that negative words attract more attention away from concurrent cognitive processes than do positive words (Pratto, 1994; Pratto & John, 1991; Wentura & Rothermund, 2003; Wentura, Rothermund, & Bak, 2000), the results of Buchner et al. also support explanations of the irrelevant speech effect within working memory models that do specify an explicit role of attention in the maintenance of information for immediate serial recall.

For three of the working memory models considered here—the feature model, Cowan's (1995) model, and one possible variant of the modular working memory model—we had assumed that the identity of to-be-recalled items may be reconstructed by successfully matching degraded short-term memory traces against a long-term memory representation, a process represented by parameter r_j in Schweickert's (1993) process model of immediate recall. In other words, long-term memory processes may be involved in immediate serial recall performance. Does this possible involvement of long-term memory in task performance imply that immediate recall experiments such as the present one cannot be used to test models of working memory? In answering this question, it is important to note that the detrimental effects of irrelevant speech in the present experiment as well as in other experiments that are typical of the field occur during rehearsal, and not during recall when, presumably, long-term memory processes come into play. Thus, immediate recall experiments such as the present one seem to represent a valid experimental paradigm for testing short-term memory models.

Aside from these considerations, the results reported here confirm the validity of Schweickert's (1993) process model of immediate recall, and they show how to obtain unique parameter estimates, correct goodness-of-fit statistics, and appropriate degrees of freedom for this prima facie nonidentifiable model so that it can be used for statistical tests that are closer to the processes assumed to underlie overt recall performance.

REFERENCES

BADDELEY, A. D. (1986). *Working memory*. Oxford: Oxford University Press, Clarendon Press.

BADDELEY, A. D. (1996). The concept of working memory. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 1-27). Hove, U.K.: Psychology Press.

BADDELEY, A. D. (2000). The phonological loop and the irrelevant speech effect: Some comments on Neath (2000). *Psychonomic Bulletin & Review*, **7**, 544-549.

BADDELEY, A. D., & LOGIE, R. H. (1999). Working memory: The multiple-component model. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 28-61). New York: Cambridge University Press.

BAMBER, D., & VAN SANTEN, J. P. H. (1985). How many parameters can a model have and still be testable? *Journal of Mathematical Psychology*, **29**, 443-473.

BAMBER, D., & VAN SANTEN, J. P. H. (2000). How to assess a model's testability and identifiability. *Journal of Mathematical Psychology*, **44**, 20-40.

BATCHELDER, W. H., & RIEFER, D. M. (1999). Theoretical and empirical review of multinomial process tree modeling. *Psychonomic Bulletin & Review*, **6**, 57-86.

BUCHNER, A., FAUL, F., & ERDFELDER, E. (1996). *G•Power: A priori, post-hoc, and compromise power analyses for the Macintosh (Version 2.1.2)* [Computer program]. Düsseldorf: Heinrich-Heine-Universität (Available at http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/).

BUCHNER, A., IRMEN, L., & ERDFELDER, E. (1996). On the irrelevance of semantic information for the "irrelevant speech" effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **49A**, 765-779.

BUCHNER, A., ROTHERMUND, K., WENTURA, D., & MEHL, B. (2004). Valence of distractor words increases the effects of irrelevant speech on serial recall. *Memory & Cognition*, **32**, 722-731.

CENTRE FOR LEXICAL INFORMATION (1991). *The CELEX lexical database. German Version 2.0* [CD-ROM]. Nijmegen: Centre for Lexical Information, Max Planck Institute for Psycholinguistics.

COHEN, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). Hillsdale, NJ: Erlbaum.

COLLE, H. A., & WELSH, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning & Verbal Behavior*, **15**, 17-32.

COWAN, N. (1995). *Attention and memory: An integrated framework*. New York: Oxford University Press.

COWAN, N. (1999). An embedded-processes model of working memory. In A. Miyake & P. Shah (Eds.), *Models of working memory: Mechanisms of active maintenance and executive control* (pp. 62-101). New York: Cambridge University Press.

CROWTHER, C. S., BATCHELDER, W. H., & HU, X. (1995). A measurement-theoretic analysis of the fuzzy logic model of perception. *Psychological Review*, **102**, 396-408.

ELLIOTT, E. M. (2002). The irrelevant-speech effect and children: Theoretical implications of developmental change. *Memory & Cognition*, **30**, 478-487.

ERDFELDER, E., FAUL, F., & BUCHNER, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, **28**, 1-11.

HU, X., & BATCHELDER, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, **59**, 21-48.

HULME, C., ROODENRYS, S., SCHWEICKERT, R., BROWN, G. D. A., MARTIN, S., & STUART, G. (1997). Word-frequency effects on short-term memory tasks: Evidence for a redintegration process in immediate serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **23**, 1217-1232.

JONES, D. M. (1993). Objects, streams, and threads of auditory attention. In A. D. Baddeley & L. Weiskrantz (Eds.), *Attention, selection, awareness, and control: A tribute to Donald Broadbent* (pp. 87-104). Oxford: Oxford University Press.

JONES, D. M., BEAMAN, C. P., & MACKEN, W. J. (1996). The object-oriented episodic record model. In S. E. Gathercole (Ed.), *Models of short-term memory* (pp. 209-237). Hove, U.K.: Psychology Press.

JONES, D. M., & MACKEN, W. J. (1993). Irrelevant tones produce an irrelevant speech effect: Implications for phonological coding in working memory. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **19**, 369-381.

JONES, D. M., & TREMBLAY, S. (2000). Interference in memory by process or content? A reply to Neath (2000). *Psychonomic Bulletin & Review*, **7**, 550-558.

LECOMPTE, D. C., & SHAIBE, D. M. (1997). On the irrelevance of phonological similarity to the irrelevant speech effect. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, **50A**, 100-118.

LUCE, R. D. (1959). *Individual choice behavior*. Oxford: Wiley.

NAIRNE, J. S. (1990). A feature model of immediate memory. *Memory & Cognition*, **18**, 251-269.

NEATH, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, **7**, 403-423.

NEELY, C. B., & LECOMPTE, D. C. (1999). The importance of semantic similarity to the irrelevant speech effect. *Memory & Cognition*, **27**, 37-44.

PRATTO, F. (1994). Consciousness and automatic evaluation. In P. M.

Niedenthal & S. Kitayama (Eds.), *The heart's eye: Emotional influences in perception and attention* (pp. 115-143). San Diego: Academic Press.

PRATTO, F., & JOHN, O. P. (1991). Automatic vigilance: The attention-grabbing power of negative social information. *Journal of Personality & Social Psychology*, **61**, 380-391.

RASCH, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press.

RIEFER, D. M., & BATCHELDER, W. H. (1988). Multinomial modeling and the measurement of cognitive processes. *Psychological Review*, **95**, 318-339.

ROBERTS, F. S. (1979). *Measurement theory with applications to decisionmaking, utility, and the social sciences*. Reading, MA: Addison-Wesley.

ROTHKEGEL, R. (1999). AppleTree: A multinomial processing tree modeling program for Macintosh computers. *Behavior Research Methods, Instruments, & Computers*, **31**, 696-700.

SALAMÉ, P., & BADDELEY, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning & Verbal Behavior*, **21**, 150-164.

SALAMÉ, P., & BADDELEY, A. D. (1989). Effects of background music on phonological short-term memory. *Quarterly Journal of Experimental Psychology*, **41A**, 107-122.

SCHWEICKERT, R. (1993). A multinomial processing tree model for degradation and redintegration in immediate recall. *Memory & Cognition*, **21**, 168-175.

SCHWEICKERT, R., CHEN, S., & POIRIER, M. (1999). Redintegration and the useful lifetime of the verbal memory representation. *International Journal of Psychology*, **34**, 447-453.

WENTURA, D., & ROTHERMUND, K. (2003). The "meddling-in" of affective information: A general model of automatic evaluation effects. In J. Musch & K. C. Klauer (Eds.), *The psychology of evaluation: Affective processes in cognition and emotion* (pp. 51-86). Mahwah, NJ: Erlbaum.

WENTURA, D., ROTHERMUND, K., & BAK, P. (2000). Automatic vigilance: The attention-grabbing power of approach- and avoidance-related social information. *Journal of Personality & Social Psychology*, **78**, 1024-1037.

NOTES

1. All model tests were performed by computing the log-likelihood goodness-of-fit statistic G^2, equivalent to the power-divergence statistic PD($\lambda = 0$), using AppleTree (Rothkegel, 1999).

2. All tests of specific model-based statistical hypotheses reported in this article follow this rationale.

3. The power calculations were conducted using the G·Power program (Buchner, Faul, & Erdfelder, 1996; Erdfelder, Faul, & Buchner, 1996).

4. In the present analyses, all multivariate test criteria correspond to the same (exact) F statistic, which is reported.

APPENDIX A

Table A1
Observed Frequencies of Correct Recall in Experiment 3 of Hulme et al. (1997)

	Serial Position						
Frequency	1	2	3	4	5	6	7
Short-Length Words							
High	461	412	372	337	284	279	406
Low	465	411	358	278	220	227	367
Medium-Length Words							
High	449	365	304	272	198	237	388
Low	428	348	255	194	131	169	355

Note—Total possible correct score in each cell is 500.

Table A2
Observed Frequencies of Correct Recall in the Present Experiment

	Serial Position						
Frequency	1	2	3	4	5	6	7
Silence							
High	399	318	260	209	120	89	97
Low	390	303	220	158	90	50	59
High-Frequency Distractors							
High	357	265	209	142	69	56	83
Low	347	250	186	115	75	55	79
Low-Frequency Distractors							
High	331	228	161	122	71	41	77
Low	322	200	138	93	58	35	71

Note—Total possible correct score in each cell is 432.

APPENDIX A (Continued)

Table A3
Words Used in the Present Experiment, Their Relative Frequencies in the Corpus Covered by the CELEX database (Centre for Lexical Information, 1991), and Their Rated Concreteness ($N = 127$)

Set 1	High-Frequency Words		Set 2	High-Frequency Words	
	Frequency	Concreteness		Frequency	Concreteness
Beginn	711	2.3	Arbeit	3,192	3.1
Gesetz	1,039	2.7	Erfolg	1,705	3.0
Gruppe	1,195	3.6	Gebiet	1,790	2.6
Nation	662	3.1	Jugend	850	3.4
Partei	2,686	3.0	Kirche	1,267	3.7
Person	898	3.2	Schule	1,098	3.6
Stimme	1,266	3.3	Stelle	1,590	2.3
Stunde	1,894	2.5	Umsatz	913	2.0
Mean	*1,295*	*3.0*	*Mean*	*1,552*	*3.0*

Set 3	Low-Frequency Words		Set 4	Low-Frequency Words	
	Frequency	Concreteness		Frequency	Concreteness
Bambus	0	3.7	Abluft	0	2.3
Berber	0	2.7	Glucke	0	3.6
Muslim	3	3.2	Mistel	0	3.4
Raspel	0	3.2	Planke	3	3.0
Seekuh	1	3.4	Schote	2	3.1
Sekret	0	3.0	Tiegel	4	2.6
Tonarm	0	2.5	Tränke	0	3.2
Zensor	3	2.1	Urmeer	0	2.3
Mean	*1*	*3.0*	*Mean*	*1*	*2.9*

Note—Average values are presented in italics below each word set. Values of 0 represent the fact that a particular word was a proper German word as defined within a standard dictionary but did not occur in the evaluated text body of about 6 million words.

APPENDIX B
Handling Identifiability Problems in Schweickert's (1993)
Multinomial Model of Immediate Serial Recall

For one serial recall position, Schweickert's (1993) model explains a single correct recall probability by two parameters, i_s and r_j (see Figure 1). Because identifiable models cannot have more parameters than degrees of freedom in the data (Bamber & van Santen, 1985, 2000; Hu & Batchelder, 1994), the model is obviously not identified for a single recall probability. Using a two-factorial $S \times J$ design where one factor (e.g., serial position) affects only the probability of an intact item representation i_s, $s = 1, \ldots, S$, and the other factor (e.g., word frequency) affects only the probability of a successful redintegration of a degraded item representation, r_j, $j = 1, \ldots, J$, it is possible to obtain additional degrees of freedom in the data that exceed or at least are equal to the number of parameters to be estimated. Unfortunately, however, this still does not suffice to obtain an identifiable model in the present case. To illustrate, let us assume that $S = J = 2$, so that a 2×2 design with four independent correct recall probabilities p_{11}, p_{12}, p_{21}, and p_{22} is available for estimating four model parameters i_1, i_2, r_1, and r_2. Although the "parameter counting rule" (Bamber & van Santen, 1985, 2000) is not violated in this case, it is possible to find for any set i_s and r_j of parameter values a different set i'_s and r'_j of parameter values predicting exactly the same recall probabilities. This also holds if $S > 2$ or $J > 2$.

PROOF: Let c be any positive real-valued constant.[B1] Define $i'_s := 1 - c(1 - i_s)$ and $r'_j := 1 - (1/c)(1 - r_j)$. Then it is easy to see that for all levels s, $s = 1, \ldots, S$, of the first factor and j, $j = 1, \ldots, J$, of the second factor, the parameters i'_s and r'_j predict the same correct recall probabilities as the parameters i_s and r_j:

$$p'_{sj} = i'_s + (1 - i'_s) r'_j$$
$$= [1 - c(1 - i_s)] + c(1 - i_s) [1 - (1/c)(1 - r_j)]$$
$$= 1 - c(1/c) (1 - i_s)(1 - r_j)$$
$$= 1 - (1 - i_s)(1 - r_j)$$
$$= i_s + (1 - i_s) r_j$$
$$= p_{sj} \text{ q.e.d.}$$

APPENDIX B (Continued)

Hence, as in the Rasch model (Rasch, 1960), in the BTL model (Luce, 1959), and in the Fuzzy Logic Model of Perception (see Crowther, Batchelder, & Hu, 1995), parameters are not uniquely determined by the cell probabilities although the number of parameters does not exceed the number of independent cell probabilities for the data. As a consequence, parameter estimates cannot be interpreted as probability estimates because probabilities require absolute scales (see also Crowther et al., 1995). Fortunately, in the present case, the parameters form a special type of interval scale where both the unit of measurement and the zero point depend on the choice of the constant c. This implies that differences between parameters of one parameter family (i.e., the r or the i parameters) form a ratio scale such that ratios of parameter differences are uniquely determined by the data.

For example, if the parameters for three successive serial positions are $i_1 = .60$, $i_2 = .30$, and $i_3 = .15$, it would not be meaningful (in the sense defined by Roberts, 1979) to interpret these numbers as probabilities of there being intact memory traces. The parameters $i_1 = .80$, $i_2 = .65$, and $i_3 = .575$, among others, could also predict the same data if the r_j parameters were modified accordingly. However, since the differences between parameters form a ratio scale, any permissible rescaling of the parameters [i.e., any transformation of the type $i_s' := 1 - c(1 - i_s)$ and $r_j' := 1 - (1/c)(1 - r_j)$] does not change the difference ratio. Thus, it remains true for all possible sets of parameters applied to our example that the difference between i_1 and i_2 is twice as large as the difference between i_2 and i_3. Thus, although the absolute values of the parameters cannot be interpreted as probabilities, we can derive meaningful statements about parameter differences and, as a corollary, about the rank order of parameters in Schweickert's (1993) model. This appears to be sufficient for most applications of the model.

The lack of identifiablity of Schweickert's (1993) model also has some consequences for goodness-of-fit tests of this model. These tests require identifiable models. One possible way to transform Schweickert's model into an identifiable model would be to design a special experimental condition that allows a priori determination one or more of the r_j or i_s parameters. For example, one could try to find an experimental condition in which the probability of correct recall equals zero.[B2] For such a condition, the parameters $i_0 = r_0 = 0$ are implied and the only possible scale factor is $c = 1$ because all other values of c would cause at least one parameter value to be outside the interval (0, 1). Unfortunately, however, this solution of the identifiability problem does not offer a solution to the goodness-of-fit test problem. The reason is that the G^2 χ^2 goodness-of-fit statistic and other possible χ^2 statistics can be shown to be asymptotically χ^2 distributed only if (1) the model is locally identifiable and (2) none of the parameter values approaches the boundary of the parameter space (e.g., Hu & Batchelder, 1994). Hence, the typical asymptotic goodness-of-fit test cannot be applied when there is one condition where both parameters lie at the boundary of the parameter space (i.e., $i_0 = r_0 = 0$).

If, as in the present case, an experimental condition allowing to determine at least one nonboundary parameter $0 < i_s < 1$ or $0 < r_j < 1$ a priori is lacking, using an arbitrary "identifiability constraint" cannot be avoided. In the case of a two-factorial design, a single parameter fixation suffices to render the model identifiable. However, because the parameters are constrained to be elements of the unit interval (0, 1), it is important to select the identifiability constraint such that no additional parameter estimate approaches the boundary of the parameter space. Roughly speaking, identifiability constraints are appropriate if they do not cause an increase in the goodness-of-fit statistic, compared with the nonidentifiable model without the identifiability constraint(s).

If multiple two-factorial designs are analyzed simultaneously, as in the case of the data reported by Hulme et al. (1997) and in the present article, one identifiability constraint is required for each two-factorial design. Because one can freely select, subject to some inequality constraints, the value of one of the parameters for each two-factorial design ahead of time, estimating this parameter does not use up a degree of freedom. Hence, the degrees of freedom of the global model test amount to the total number of independent data categories minus the number of model parameters plus the number of parameters that need to be constrained in order to render the model identifiable.

This was the strategy we employed in the analyses reported in the present article.

NOTES

B1. As noted by an anonymous reviewer, although this proof holds for any positive real-valued c, in practice only c values not too different from 1 will be relevant. Both c values much larger than 1 and c values close to zero would cause scale transformations resulting in parameter values outside the permissible interval (0, 1).

B2. The authors thank R. Schweickert for suggesting this possibility.