



# Disruption of order information by irrelevant items: A serial recognition paradigm

Jens Gisselgård<sup>a,\*</sup>, Julia Uddén<sup>a</sup>, Martin Ingvar<sup>a</sup>,  
Karl Magnus Petersson<sup>a,b,c</sup>

<sup>a</sup> *Cognitive Neurophysiology Research Group, Department of Clinical Neuroscience, Karolinska Institutet, MRC N8, Karolinska Hospital, S-171 76 Stockholm, Sweden*

<sup>b</sup> *F. C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, The Netherlands*

<sup>c</sup> *CSI, Centre for Intelligent Systems, University of Algarve, Faro, Portugal*

Received 4 September 2005; received in revised form 12 April 2006; accepted 14 April 2006

Available online 14 June 2006

---

## Abstract

Irrelevant speech effect (ISE) is defined as a decrement in visually presented digit-list short-term memory performance due to exposure to irrelevant auditory material. Perhaps the most successful theoretical explanation of the effect is the *changing state hypothesis*. This hypothesis explains the effect in terms of confusion between amodal serial order cues, and represents a view based on the interference caused by the processing of similar order information of the visual and auditory materials. An alternative view suggests that the interference occurs as a consequence of the similarity between the visual and auditory contents of the stimuli. An important argument for the former view is the observation that ISE is almost exclusively observed in tasks that require memory for serial order. However, most short-term memory tasks require that both item and order information be retained in memory. An ideal task to investigate the sensitivity of maintenance of serial order to irrelevant speech would be one that calls upon order information but not item information. One task that is particularly suited to address this issue is serial recognition. In a typical serial recognition task, a list of items is presented and then probed by the same list in which the order of two adjacent items has been transposed. Due to the re-presentation of the encoding string, serial recognition requires primarily the serial order to be maintained while the content of the presented items is deemphasized. In demonstrating a highly significant ISE of changing versus steady-state auditory items in a serial recognition task, the present finding lends support for and extends previous empirical findings

---

\* Corresponding author. Tel.: +46 8 51776116; fax: +46 8 51773266.  
E-mail address: [jens.gisselgard@ki.se](mailto:jens.gisselgard@ki.se) (J. Gisselgård).

suggesting that irrelevant speech has the potential to interfere with the coding of the order of the items to be memorized.

© 2006 Elsevier B.V. All rights reserved.

*PsychINFO classification:* 2326

*Keywords:* Irrelevant speech; Irrelevant sound; Short-term memory; Selective attention; Recognition

---

## 1. Introduction

Under certain circumstances, information that is completely irrelevant and that we are actively trying to ignore still severely disrupts our cognitive performance. A well known example of this phenomenon is the irrelevant speech effect (Colle & Welsh, 1976; Jones & Macken, 1993; LeCompte, 1994; Neath, 2000; Salame & Baddeley, 1982; for reviews, see Banbury, Tremblay, Macken, & Jones, 2001; Jones & Morris, 1992). In a typical irrelevant speech experiment, lists of visually presented items are presented for immediate serial recall, while the subjects are exposed to irrelevant auditory material. In attempting to explore the cause of the effect, previous studies have emphasized either the nature of the irrelevant sounds or the characteristics of the material to be memorized. The latter approach involves a careful examination of which tasks produce an effect and which tasks do not. While the task most frequently employed is a standard serial recall task (e.g. Colle & Welsh, 1976; Ellermeier & Zimmer, 1997; LeCompte, 1995; Jones, Madden, & Miles, 1992; Salame & Baddeley, 1982), the effect has been investigated in a range of other tasks, such as free recall (Beaman & Jones, 1998; LeCompte, 1994; Richardson, 1984; Salame & Baddeley, 1990), paired associate cued recall (Beaman & Jones, 1997; LeCompte, 1994), a missing item task (Beaman & Jones, 1997), and probe recall or recognition tasks (Beaman & Jones, 1997; Henson, Hartley, Burgess, Hitch, & Flude, 2003; LeCompte, 1994). The pattern of results obtained from these studies is rather mixed and is consistent with the notion that when serial rehearsal is employed by the subjects, an irrelevant speech effect is observed (Beaman & Jones, 1997; Jones & Macken, 1993; Salame & Baddeley, 1990). Since the effect occurs primarily in tasks that require, or is subject of, maintenance of serial order information, it has been suggested that the interference is mainly one of confusion between different cues to serial order, the so called *changing state hypothesis* (Jones, 1993; Jones, Macken, & Murray, 1993; Jones et al., 1992). Although the precise mechanism behind the interference is not yet well understood, the argument put forward is that the interference is based on the disruption of the order of events, so that amodal order cues from the relevant as well as the irrelevant stream of information may conflict with each other. A comparison of results from the various tasks employed is complicated by the fact that they often have additional properties in common other than seriality. In particular, most tasks require that both item and order information be retained in memory. An ideal task to investigate the sensitivity of maintenance of serial order to irrelevant speech would be one that calls upon order information but not item information. One task approaching this ideal is serial recognition, also known as matching span (Allport, 1984; Gathercole, Service, Hitch, Adams, & Martin, 1999; but see Farrand & Jones, 1996, for an alternative approach). Generally speaking, this task has several potential benefits when compared to a standard serial recall task. In a typical serial recognition task, a list of items is presented and then probed by the same list in which

the order of two adjacent items has been transposed. Due to the re-presentation of the encoding string, this task only claims memory for order while deemphasizing item information, whereas in a serial recall both item and order information is required. It is also the case that this task does not depend on the preparation and organization of articulatory output compared to a spoken serial recall task (Gathercole, Pickering, Hall, & Peaker, 2001). As a consequence, it has proved to be useful when assessing the phonological storage capacity with individuals with language and speech-motor inabilities (Allport, 1984; Campbell & Butterworth, 1985; Jarrold, Baddeley, & Phillips, 2002; Martin & Breedin, 1992; Shallice & Warrington, 1977). An interesting consequence of this field of application is that serial recognition can prove to be a useful test of phonological short-term memory in experimental environments in which oral response modes are not applicable, such as in an fMRI setting (for example, see Gisselgård, Uddén, Fransson, Ingvar & Petersson, in preparation; Henson, Burgess, & Frith, 2000). A third reason for using serial recognition rather than the more conventional serial recall task is that it is suggested to rely to a lesser extent on long-term knowledge. Considering that long-term memory knowledge, such as lexicality, can contribute substantially to serial recall performance implicates that scores on conventional verbal short-term memory tasks reflect at least in part the interplay between both of these memory systems. Noticeably, a work by Gathercole et al. (2001) shows that when serial recall for words and non-words is studied, a large word advantage occurs, but disappears when it is tested by serial recognition. Similarly, Thorn, Gathercole, and Frankish (2002) find a large first-language advantage on bilingual serial recall performance, while the recognition performance was comparable in the two languages. Apart from these differences, both serial recall and serial recognition are sensitive to vocabulary knowledge and phonological similarity, suggesting that both tasks reflect a common temporary storage of phonological information (Gathercole et al., 2001). In addition, serial recognition scores correlate highly with scores in serial recall of non-words (Gathercole et al., 1999), also suggesting that they reflect the same phonological short-term capacity. In summary, unlike recall, serial recognition provides a relatively pure estimate of phonological short-term memory capacity. While several previous studies examine irrelevant speech effects in experiments using standard recognition procedures, very few studies have used the serial recognition paradigm together with irrelevant speech. Henson et al. (2003, Experiment 1) investigated irrelevant speech with this task, but did not examine the effects of changing state items compared to steady-state items (cf., the Discussion section). The present study allows a direct comparison between changing state and steady-state sounds within the serial recognition paradigm. An effect of changing state sounds vs. steady-state sounds would be consistent with the changing state prediction, lending support for the claim that changing-state speech has the potential to interfere with the order of items maintained in short-term memory.

## 2. Materials and methods

### 2.1. Participants

Sixty-two undergraduate volunteers (28 female; mean age 24 [SD = 3, range 18–32]) with a median of 1.5 (range 0–5) years of university level education participated in the study. All subjects were naive as to the literature on the irrelevant speech effect and to the hypothesis being investigated in the study. They were not paid for participating, but they were informed that a subset of participants would be selected and asked to participate

in a re-test session in the MR-scanner (Gisselgård, Uddén, Fransson, Ingvar & Petersson, in preparation). The subjects were pre-screened and none used any medication, nor had a history of drug abuse (including nicotine), head trauma, neurological or major psychiatric illness, or family history of neurological or psychiatric illness. The study was approved by the local Ethics Committee at the Karolinska Hospital (02-361). Informed consent was obtained from all the subjects.

## 2.2. Stimuli

In a within subject factorial design, the effects of two levels of irrelevant speech and three levels of working memory load were investigated.

### 2.2.1. Visual stimuli

Lists of 6, 8, or 10 digits were generated by random selection without replacement from the digits 0–9. The visual stimuli were presented with Presentation software v. 0.81 (<http://nbs.neuro-bs.com>) and consisted of lists of 6 ('low load'), 8 ('medium load') or 10 ('high load') digits (black text, Tahoma, size 60 on white background) presented in a randomized order. Half of the lists were followed by a positive probe and half by a negative probe in which two adjacent digits were randomly transposed (no extreme positioned pairs were transposed). Working memory load is defined as the number of items per time unit. In order to have constant presentation times across list lengths, inter-stimulus intervals varied between conditions. Stimulus offset was set to 850 ms for low load lists, 533 ms for medium load and 350 ms for high memory load lists, giving a total presentation time of 6 s (6050 ms, 6131 ms and 6150 ms for low, medium and high load lists, respectively<sup>1</sup>) across all list lengths. The onset of the digits was 300 ms in all conditions.

### 2.2.2. Auditory stimuli

The irrelevant speech was present at two levels, a single and a multiple item level. The single item condition (S) consisted of repetitions of the CV-pseudo-word "da" [da:], while in the multiple item condition (M), seven phonologically dissimilar CV-syllables (ne [ne], li [li], to [tu:], vu [vu], py [py], bå [bɔ:], nö [nœ:]) were presented in a randomized order (although no repeats were allowed). The auditory items were digitally recorded at 44 kHz 16-bit mono, with a male voice at an approximately even pitch and subsequently edited in Sony Sound Forge 7.0 (<http://www.sony.com/mediasoftware>). The multiple items were digitally shifted in pitch with one semitone between items, giving each item a unique pitch in an approximate semitone scale with seven tones. Both single and multiple items were presented at a frequency of 100 items/min. Each list of digits was associated with either the single item or the multiple item level of irrelevant speech. The speech conditions as well as the number of correct and incorrect lists were counterbalanced over list lengths.

## 2.3. Procedure

In each of six experimental blocks, 24 encoding-delay-recognition cycles were presented in a randomized order. After the word "Ready" was displayed, a list of digits

---

<sup>1</sup> The precision is constrained by the monitor refresh rate (60 Hz).

was sequentially presented in the centre of a 17 in. computer screen. Subsequently, a delay of 4 s was indicated by a symbol (\*), and during the following 4 s recognition interval, a complete new string of digits was presented. The first string of digits was presented serially, whereas the recognition string was presented simultaneously. During the recognition interval, the subjects were prompted to respond as “correct” or “incorrect” by clicking on the left or right mouse button, respectively. Reaction times were collected with an external standard mouse device that did not move. This method is considered to be accurate although it can produce a small and stable delay ( $31 \pm 2$  ms) (Segalowitz & Graves, 1990). The next list appeared 3 s after the recognition period, independent of whether the subjects had responded or not, in a paced paradigm. The subjects were encouraged to answer both as correctly and as quickly as possible, to guess if they did not know the correct response and were informed that only the first response in each list would be scored. They were also told that negative probes would only differ in the order of two adjacent items. The irrelevant speech was presented through headphones at a loud but not uncomfortable level<sup>2</sup> during the full presentation-delay-probe cycle of each trial. The subjects were instructed to ignore any sounds presented in the headphones and to memorize the digits in groups of two or three by subvocal rehearsal. Overt vocal rehearsal was discouraged. After two short (12 lists in each) trial blocks, the subjects commended with the six experimental blocks. These lasted for 6 min 50 s each. Between each experimental block subjects were allowed a short break of about a minute before they engaged in the next experimental block. The complete session took about 45 min. After the test, the subjects were asked about their mnemonic strategy and the perceived difficulty of each memory load as well as the subjective susceptibility to the irrelevant sounds.

### 3. Results

#### 3.1. Accuracy

Performance was measured as the percentage of the correct responses in correct as well as incorrect lists. On average, the proportions of correct responses (chance level = 0.5) were 80% (SD = 9) for the lists presented with a single item background and 77% (SD = 8) for the lists presented with a multiple items background. The distribution of individual effect sizes, as measured by *S*–*M* scores, is displayed in Fig. 1.

The probability correct for the different levels of working memory load was 91% (SD = 7), 78% (SD = 10) and 66% (SD = 9), on the low, medium and high load, respectively. The probability correct for the different list types was 80% (SD = 9) for correct lists (positive probes) and 76% (SD = 11) for incorrect lists (negative probes). The data were approximately normally distributed. The experimental design was  $2 \times 2 \times 3$  factorial, with two levels of list type (positive or negative probes), two levels of irrelevant speech (changing or steady-state), and three levels of working memory load (6, 8 or 10 items). A three-way repeated measures ANOVA revealed a significant irrelevant speech effect [ $F(1, 61) = 14.6$ ,  $MS_e = 0.02$ ,  $p < 0.001$ ] and a significant main effect of working memory

<sup>2</sup> Since the intensity of sound has been previously shown not to interact with the irrelevant speech effect within the range of a whisper to a shout (Colle & Welsh, 1976; Ellermeier & Hellbrueck, 1998), the volume was on some occasions adjusted according to the subjects' preference. A condition was that it should be fairly loud but that the level itself should not be a nuisance.

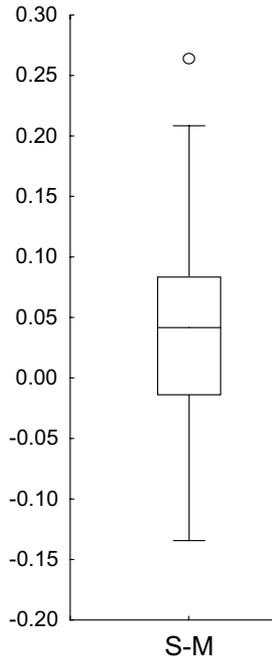


Fig. 1. Box-plot of the distribution of individual irrelevant speech effects as measured by S–M scores.

load [ $F(2, 122) = 266.6$ ,  $MS_e = 0.01$ ,  $p < 0.001$ ], as well a significant main effect of list type [ $F(1, 61) = 4.8$ ,  $MS_e = 0.06$ ,  $p < 0.05$ ]. None of the interactions reached significance ( $F < 1.2$ ). A sphericity test was performed on the ANOVA. With the exception of a type  $\times$  load interaction, none of the other terms in this test reached significance. Thus, it was concluded that the sphericity assumption was satisfactorily met. The varied speech exerted a decrement in performance that was approximately evenly distributed, although slightly diminished, across working memory loads (Fig. 2a). Planned comparisons confirmed that there was a significant simple main effect of irrelevant speech at the low [ $F(1, 61) = 21.4$ ,  $MS_e = 0.008$ ,  $p < 0.001$ ] and medium loads [ $F(1, 61) = 5.72$ ,  $MS_e = 0.02$ ,  $p < 0.05$ ], but not at the high load ( $p = 0.20$ ). Analysis of these and subsequent performance data using arcsin transformations or a guessing correction of hits minus false alarms gave equivalent results. In addition, a signal detection analysis was performed, in which a discrimination index  $d'$  is calculated as a measure of performance which is independent of response bias (Hochhaus, 1972). The discrimination index is the distance between the mean hit-rate and the false alarm-rate in units of standard deviations. A two-way repeated measures ANOVA (irrelevant speech  $\times$  working memory load) of the  $d'$ -data revealed equivalent results with the accuracy analysis based on percent correct. A two-way repeated measures ANOVA performed on the bias-data was non-significant.

In order to examine the relationship between individual working memory capacity scores and effects of irrelevant speech, linear correlations (Pearson  $r$ ) were calculated between performance scores (mean percent correct) and irrelevant speech effects (single vs. multiple items) at each level of working memory load. Although weak correlations

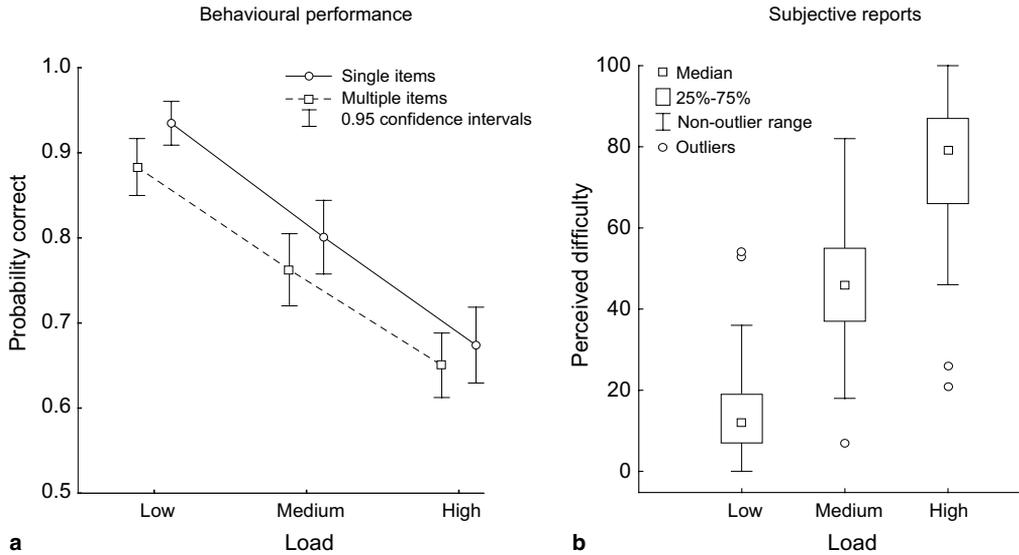


Fig. 2. (a) The probability of correct responses in the presence of single and multiple auditory items and (b) a box-plot of the perceived difficulty across the low, medium and high levels of working memory load.

were observed at the low ( $r = -0.29$ ,  $p < 0.05$ ) and high ( $r = 0.16$ , n.s.) working memory loads, the correlation at the medium load was close to zero ( $r = 0.02$ , n.s.).

### 3.2. Speed

The mean reaction times (RT) for each condition are reported in Table 1. A repeated measures ANOVA showed a significant main effect of load [ $F(2, 122) = 63.8$ ,  $MS_e = 0.13$ ,  $p < 0.001$ ] and sound [ $F(1, 61) = 19.0$ ,  $MS_e = 0.04$ ,  $p < 0.001$ ], but not of list type [ $F(1, 61) = 1.35$ ,  $MS_e = 0.09$ ]. There was also a significant interaction between sound and load [ $F(2, 122) = 10.2$ ,  $MS_e = 0.05$ ,  $p < 0.001$ ], while the remaining interactions were not significant. Planned comparisons showed that subjects responded significantly faster during the single item auditory background compared to the multiple item level at the low [ $F(1, 61) = 28.2$ ,  $MS_e = 0.044$ ,  $p < 0.001$ ] and medium loads [ $F(1, 61) = 7.07$ ,  $MS_e = 0.046$ ,  $p < 0.01$ ]. However, at the high working memory load, reaction times did not differ to a significant degree ( $p = 0.20$ ).

Table 1  
Mean reaction times (s) in recognition for the low, medium and high working memory load

Irrelevant stimuli	Working memory load					
	Low		Medium		High	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Single items	1.51	0.27	2.18	0.37	2.60	0.41
Multiple items	1.65	0.30	2.25	0.37	2.56	0.40

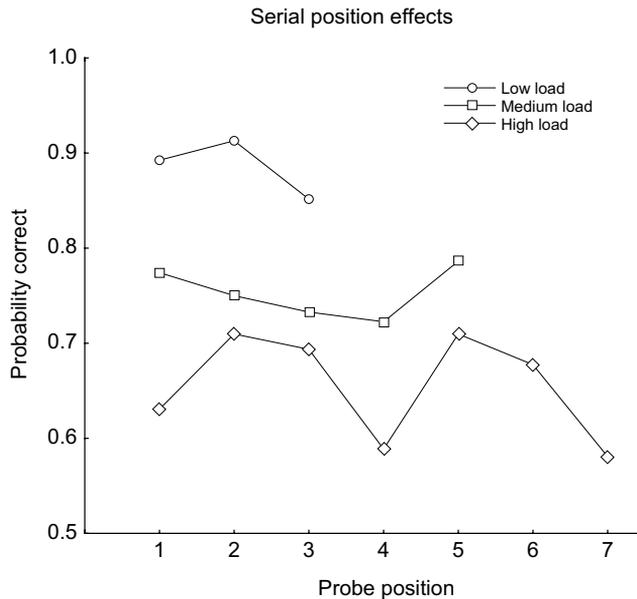


Fig. 3. Probability correct as a function of probe position for each level of working memory load. Note that since the first and last pair of digits in each string was not transposed, serial position data only reflects  $k-3$  possible transpositions, where  $k$  is the number of items in the string.

### 3.3. Serial position effects

Serial position effects on negative probes are displayed in Fig. 3. Data from each level of working memory load was subjected to a repeated measures ANOVA with the serial position as a within-factor. For the low load condition, the ANOVA was significant [ $F(2, 122) = 4.14$ ,  $MS_e = 0.06$ ,  $p < 0.05$ ]. A post-hoc test (Newman Keuls) revealed that there was a significant effect only between the second and third position ( $p < 0.05$ ). There were no significant effects of serial position in the medium [ $F(4, 244) = 1.23$ ,  $MS_e = 0.04$ ,  $p = 0.29$ ] or high load conditions [ $F(6, 366) = 1.62$ ,  $MS_e = 0.12$ ,  $p = 0.14$ ].

### 3.4. Subjective reports

In the post-experimental questionnaire, 35 subjects stated that the multiple item sound was the most disturbing, and 19 stated that it was more difficult to perform during the single item sound. Eight subjects perceived the sounds to be equally disturbing. Including this subjective measure into the above ANOVA of percent correct as a categorical predictor (factor), it came out as not significant ( $F < 1$ ). However, there was a significant interaction between this factor and irrelevant speech, that is, the magnitude of the effect of varied speech was dependant on what sound the subjects perceived was the most difficult [ $F(2, 59) = 3.98$ ,  $MS_e = 0.02$ ,  $p < 0.05$ ]. A post-hoc Tukey test revealed that this difference was significant only for those subjects that had stated the multiple item background to be the most demanding ( $p < 0.001$ ). Apparently, subjective ratings of difficulty were less reflective of the actual performance in the single item than in the multiple item condition.

No other interaction effects were significant in this analysis, while the main effects of sound ( $p < 0.05$ ), load ( $p < 0.001$ ) and type ( $p < 0.05$ ) remained significant.

The perceived difficulty was rated by the subjects on a visual analogue scale ranging from 0 to 100 mm, revealing a linear increase across the levels of working memory load (Fig. 2b). All subjects reported that they had mainly used the suggested mnemonic strategy, where digits most commonly were memorized in pairs or triplets. Twenty-five subjects specifically stated that they had also used one or several complementary strategies. These strategies could be classified as being either associative (for example, associating numbers with years of historic or personal value), trend sensitive (upward or downward tendency of the numbers), visual (some subjects stated they created an inner vision of a part of the digit sequence), or melodic (a few subjects reported that they remembered the digits in a tuneful fashion). Some subjects also reported that the digit zero was left out of the remembered sequence as a “blank space”.

#### 4. Discussion

The present experiment was designed in order to isolate the effect of varied speech attributed to a disruption of serial order information by contrasting changing state with steady-state auditory items in a serial recognition task. The results demonstrated a highly significant changing state effect of multiple vs. single auditory items. In the serial recognition paradigm, memory for the order between adjacent items is crucial, while minimal demands are put on the representation of content due to the re-presentation of the items at test. This contrasts with serial recall, which requires the storage of both item and order information. Consequently, the results of the present experiment add strong support for and extend previous empirical findings (Beaman & Jones, 1997; Henson et al., 2003; Jones & Macken, 1993), claiming that irrelevant varied speech has the potential to interfere with the coding of the order of the items to be memorized.

Although the vast majority of studies on the irrelevant speech effect employ some variation of the serial recall task, effects have been reported in recognition tasks as well. In specific, irrelevant speech effects have been found in both item probe recognition (LeCompte, 1994) as well as in list probe (i.e., ‘matching span’) recognition (Henson et al., 2003). Item probe tasks call for the recognition of a single item rather than a sequence of items, as does a list probe task. Thus, an item probe task makes no explicit demand for serial order memory and should not produce an irrelevant speech effect according to the changing state prediction. In experiments 5A–C of LeCompte’s study (1994), lists of 8, 12 and 16 items to be memorized were used. Single items or pairs of items (target plus lure) were used as probes and the irrelevant auditory background consisted of non-words, while continuous or bursts of white noise were used as a control. Contrary to the theoretical expectation, the irrelevant speech produced significant effects in the order of 5–10%. However, these results have failed to replicate in subsequent experiments in which the subjects’ rehearsal strategies were taken into consideration (Beaman & Jones, 1997). Beaman and Jones argue that even though the recognition tasks employed by LeCompte make no explicit demands for the maintenance of serial order, subjects may still have used a strategy based on serial rehearsal. Following this line of reasoning, the recognition task in LeCompte’s study shares the serial rehearsal component as the serial recognition task used in the present study. A common procedure when analysing data from serial recall tasks is to produce serial position curves. A linear slope with primacy and recency effects is thought to reflect

the serial processing of the to-be remembered items. Taken together, the results from the present serial position analysis show some evidence of interaction between accuracy and position. The line plot at the medium level of working memory load bears some resemblance to the serial position curve typically found in serial recall tasks (see Fig. 3). However, the pattern is rather mixed for the different working memory loads, probably in part reflecting floor effects at the medium and high levels of working memory load. Considering also the limited range of possible transitions, the results should be treated with caution.

In a study by Henson et al. (2003, Experiment 1), a task called 'list probe' was used that was similar to the one in the present study. Lists of 5, 6, or 7 letters were followed by either positive or negative list-probes. The irrelevant speech consisted of spoken sentences. An item probe task served as a control task in which a single positive or negative probe appeared, and thus the serial order of items was not explicitly required. Apart from a small, non-significant effect, the item probe task did not produce a significant effect of irrelevant speech, consistent with the results of Beaman and Jones (1997). The list probe task produced a moderate (8%) irrelevant speech effect compared to a quiet background. However, a consequence of using a speech vs. quiet comparison is that it is difficult to interpret what quality of the speech was interfering with the visual input. The changing state hypothesis holds that the key feature of speech which causes the interference with short-term memory is that it is changing in states. In order to be able to extend the effect as explained by the changing state hypothesis to the serial recognition task, an effect has to persist when contrasting changing with steady-state irrelevant auditory items.

The observed effect of varied speech was relatively small (4%). Considering the closely matched auditory conditions, however, this effect possibly reflects a disruption in short-term memory for serial order brought about by the phonological and/or pitch variability in the multiple auditory stream of information. In comparison, unspecific speech vs. quiet comparisons typically produce smaller effects. For example, there is a minor effect in the item probe task of Henson et al.'s study (2003), consistent with similar findings of irrelevant speech effects in tasks where serial order is not required and in which a quiet control is used (Beaman & Jones, 1997). In keeping with this observation, there is typically a small task unspecific effect of steady-state sounds as well. Notably, in both Experiments 3 and 4 of Jones et al.'s study (1992), the effect of steady-state sounds compared to quiet also disrupted the performance by 3% and in Experiment 4, this effect was significant. This and similar observations (Jones, 1994, Experiment 1; LeCompte, 1995, Experiment 1–4) have led to the suggestion that the irrelevant speech effect consists of two effects: one specific to tasks that require memory for serial order, and one more subtle effect that is unspecific in the sense that it affects any task requiring attention and memory (cf. LeCompte, 1996; Neath, 2000). The present results show a difference between multiple and single auditory conditions, and is therefore not a general speech versus quiet effect. Generally speaking, changing-state sounds compared to steady-state sounds in controlled laboratory settings consistently produce an irrelevant sound effect of 5–15%. In keeping with the changing state hypothesis, this effect shares the same underlying mechanisms as the considerably larger effect seen with naturalistic speech, typically producing a disruption in performance in the order of 15–30% (Colle & Welsh, 1976) or even 50% (Ellermeier & Zimmer, 1997). It remains an open question, however, whether these effects are only quantitatively different from each other.

The general increase in RT due to the working memory load observed in the present experiment is consistent with the earlier findings that RT increase with list length, approximated by a linear function (Sternberg, 1969). The best linear fit gave a slope of 250 ms/item

increased in working memory load. This is a slope close to other tasks which are thought to require serial rehearsal and too slow to merely reflect a visual search task (Henson et al., 2003). Subjects used significantly more time to respond during the multiple item sound compared to the single item sound at the low and medium working memory loads. The absolute difference was 140 ms at the low load, which is almost 10% of the total reaction time. In terms of relative RT-effects, this can be considered as a large effect. For example, the stroop task consistently produces differences between congruent and incongruent conditions of 20–40 ms or 5% of the total reaction time (Atkinson, Drysdale, & Fulham, 2003; Khateb, Michel, Pegna, Landis, & Annoni, 2000). In a cross-modality dual-task, interference of auditory choice RT task by a visual encoding task showed prolonged RT of 50 ms or 7% (Azuma, Prinz, & Koch, 2004). In a recent study by Hadlington, Bridges, and Darby (2004), prolonged reaction times due to irrelevant speech was reported. In comparison to the present study, the effect size was considerably smaller (30 ms), while both studies share mean reaction times of about two seconds. The study of Hadlington et al. used similar visual and auditory stimuli, and the task required memory of serial order of five consonants, comparable to the low working memory load conditions in the present study. It should be noted that in the study of Hadlington et al. a quiet control condition was used, while the present study used repeated single items as a baseline. The accepted figures for mean simple reaction times for visual stimuli measured with college-age individuals have been about 190 ms (Galton, 1899; Welford, 1980). Hence, it is not likely that the RT-effect in the present experiment could be explained by elevated times of planning and executing a motor response, since this would have doubled the time to execute the motor response. Furthermore, earlier RT-data from a study by Salame and Baddeley (1986) could not show that irrelevant speech extended reaction times in three different simple decision tasks, for instance, one where subjects had to decide whether two letters were the same or different. These negative findings support the suggestion that irrelevant speech does not simply elevate the time to perform simple tasks as these decision tasks, or the motor response part of a more complex task, but rather slows down higher cognitive processes. However, the small effect size from the study of Hadlington et al. does not support this view. A possible explanation for this discrepancy is that a larger effect was not captured in Hadlington et al.'s study due to ceiling effects in performance, as indicated by the low error rates reported. Conversely, floor effects in the performance data of the present study might explain the absence of a difference in RT between the speech conditions at the high level of load.

A common procedure in short-term memory experiments is to use different loads in order to allow ceiling or floor effects in overall performance levels. In the study of Henson et al. (2003), lists of 5, 6, and 7 items were used, while in the present study lists of 6, 8 and 10 items were used. The present results showed a clear effect of working memory load on the performance of the subjects, with a performance level of about 90% at the lowest load, and about 65% at the highest load. The linearity in performance data in relation to working memory load is paralleled by the subjective reports of the perceived difficulty of the task at the different levels of load (Fig. 2). Considering that the chance level is 50%, the low mean score on the high level of working memory load suggests that the working memory capacity was reaching its limits. Load and capacity are interrelated concepts, such that when load is increased, the working memory capacity reaches its limits. Previous findings show that subjects are likely to abandon phonological coding when performance rates drop below some critical level. In keeping with this hypothesis, the subjective reports indicated that the subjects engaged in a number of additional strategies besides subvocal

rehearsal. In particular, we were interested in knowing whether a clearly supra-span memory load would be associated with a decreased irrelevant speech effect. While there was a slight tendency for such a decrement, this has to be investigated further before any conclusions can be drawn. Examining the main effects of irrelevant speech and working memory load at a group level might have little to say about the relationship between individual working memory capacity and the ability to sustain irrelevant distractors. It has been suggested that while high perceptual load prevents distractor interference, high ‘executive’ load increases interference by irrelevant distractors (Lavie, 2005). This hypothesis implies that the individual working memory capacity scores of the subjects in the present study might correlate negatively with their sensitivity to irrelevant speech, such that the better they perform on the serial recognition task, the better they cope with the irrelevant speech. While there was some support for this suggestion at the low level of working memory load ( $r = -0.29$ ), the medium and high levels of load showed virtually no correlation. Lavie’s hypothesis is however based mainly on studies conducted using stimuli within a single, usually visual, modality and might not apply to the effects of irrelevant speech. A few studies have examined cross-modal effects, but the results from these studies are rather mixed. Overall, our results suggest that there is no linear relation between visual short-term memory capacity and the ability to sustain irrelevant speech, consistent with the previous findings (Beaman, 2004; Ellermeier & Zimmer, 1997; Neath, Farley, & Surprenant, 2003).

## 5. Conclusion

The results of the present experiment demonstrated a significant effect of varied irrelevant speech when comparing multiple versus single auditory items in a serial recognition task. This task has several potential benefits when compared to a standard serial recall task. Most importantly, serial recognition requires primarily that serial order is maintained in short-term memory, while minimizing the need for a representation of the content of the presented items. In demonstrating a significant effect between the degree of interference of multiple and single auditory items, a measure uncontaminated of general sound vs. quiet effects, the present finding adds to the notion of the sequential specificity of the irrelevant speech effect and extends it to the serial recognition paradigm.

## Acknowledgements

This work was funded by grants from the Swedish Medical Research Council (8246), the Karolinska Institutet, the Swedish Bank Tercentenary Foundation and the Knut and Alice Wallenberg Foundation.

We thank Dr. Tom Campbell for helpful comments on a previous draft of this paper, and Christian Forkstam for creating the presentation scenario and pcl files.

## References

- Allport, D. A. (1984). Auditory-verbal short-term memory and conduction aphasia. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 313–326). Hove, UK: Lawrence Erlbaum Associates Ltd.
- Atkinson, C. M., Drysdale, K. A., & Fulham, W. R. (2003). Event-related potentials to Stroop and reverse Stroop stimuli. *International Journal of Psychophysiology*, *47*(1), 1–21.

- Azuma, R., Prinz, W., & Koch, I. (2004). Dual task slowing and the effects of cross-task compatibility. *Quarterly Journal of Experimental Psychology A*, 57(4), 693–713.
- Banbury, S. P., Tremblay, S., Macken, W. J., & Jones, D. M. (2001). Auditory distraction and short-term memory: Phenomena and practical implications. *Human Factors*, 43(1), 12–29.
- Beaman, C. P. (2004). The irrelevant sound phenomenon revisited: what role for working memory capacity? *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 30(5), 1106–1118.
- Beaman, C. P., & Jones, D. M. (1997). Role of serial order in the irrelevant speech effect: Tests of the changing-state hypothesis. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 23(2), 459–471.
- Beaman, C. P., & Jones, D. (1998). Irrelevant sound disrupts order information in free recall as in serial recall. *Quarterly Journal of Experimental Psychology A*, 51, 615–636.
- Campbell, R., & Butterworth, B. (1985). Phonological dyslexia and dysgraphia in a highly literate subject: A developmental case with associated deficits of phonemic processing and awareness. *Quarterly Journal of Experimental Psychology: Human Experimental Psychology*, 37A(3), 435–475.
- Colle, H. A., & Welsh, A. (1976). Acoustic masking in primary memory. *Journal of Verbal Learning and Verbal Behavior*, 15, 17–32.
- Ellermeier, W., & Hellbrueck, J. (1998). Is level irrelevant in “irrelevant speech”? Effects of loudness, signal-to-noise ratio, and binaural unmasking. *Journal of Experimental Psychology: Human Perception and Performance*, 24(5), 1406–1414.
- Ellermeier, W., & Zimmer, K. (1997). Individual differences in susceptibility to the “irrelevant speech effect”. *Journal of Acoustical Society of America*, 102(4), 2191–2199.
- Farrand, P., & Jones, D. (1996). Direction of report in spatial and verbal serial short-term memory. *Quarterly Journal of Experimental Psychology A*, 49(1), 140–158.
- Galton, F. (1899). On instruments for (1) testing perception of differences of tint and for (2) determining reaction time. *Journal of the Anthropological Institute*, 19, 27–29.
- Gathercole, S. E., Pickering, S. J., Hall, M., & Peaker, S. M. (2001). Dissociable lexical and phonological influences on serial recognition and serial recall. *Quarterly Journal of Experimental Psychology A*, 54(1), 1–30.
- Gathercole, S. E., Service, E., Hitch, G. J., Adams, A.-M., & Martin, A. J. (1999). Phonological short-term memory and vocabulary development: Further evidence on the nature of the relationship. *Applied Cognitive Psychology*, 13(1), 65–77.
- Hadlington, L., Bridges, A. M., & Darby, R. J. (2004). Auditory location in the irrelevant sound effect: The effects of presenting auditory stimuli to either the left ear, right ear or both ears. *Brain & Cognition*, 55(3), 545–557.
- Henson, R. N., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia*, 38(4), 426–440.
- Henson, R., Hartley, T., Burgess, N., Hitch, G., & Flude, B. (2003). Selective interference with verbal short-term memory for serial order information: a new paradigm and tests of a timing-signal hypothesis. *Quarterly Journal of Experimental Psychology A*, 56(8), 1307–1334.
- Hochhaus, L. (1972). A table for the calculation of  $d'$  and  $\beta$ . *Psychological Bulletin*, 77(5), 375–376.
- Jarrold, C., Baddeley, A. D., & Phillips, C. E. (2002). Verbal short-term memory in Down syndrome: A problem of memory, audition, or speech? *Journal of Speech, Language, & Hearing Research*, 45(3), 531–544.
- Jones, D. (1993). Objects, streams, and threads of auditory attention. In A. Baddeley & L. Weiskrantz (Eds.), *Attention selection awareness & control* (pp. 87–104). Oxford: Clarendon Press.
- Jones, D. M. (1994). Disruption of memory for lip-read lists by irrelevant speech: further support for the changing state hypothesis. *Quarterly Journal of Experimental Psychology A*, 47(1), 143–160.
- 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(2), 369–381.
- Jones, D. M., Macken, W. J., & Murray, A. C. (1993). Disruption of visual short-term memory by changing-state auditory stimuli: the role of segmentation. *Memory & Cognition*, 21(3), 318–328.
- Jones, D. M., Madden, C., & Miles, C. (1992). Privileged access by irrelevant speech to short-term memory: The role of changing state. *Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, 4, 645–669.
- Jones, D., & Morris, N. (1992). Irrelevant speech and serial recall: implications for theories of attention and working memory. *Scandinavian Journal of Psychology*, 33(3), 212–229.
- Khateb, A., Michel, C. M., Pegna, A. J., Landis, T., & Annoni, J.-M. (2000). New insights into the Stroop effect: a spatiotemporal analysis of electric brain activity. *Neuroreport*, 11(9), 1849–1855.

- Lavie, N. (2005). Distracted and confused?: selective attention under load. *Trends in Cognitive Sciences*, 9(2), 75–82.
- LeCompte, D. C. (1994). Extending the irrelevant speech effect beyond serial recall. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 20(6), 1396–1408.
- LeCompte, D. C. (1995). An irrelevant speech effect with repeated and continuous background speech. *Psychonomic Bulletin & Review*, 2(3), 391–397.
- LeCompte, D. C. (1996). Irrelevant speech, serial rehearsal, and temporal distinctiveness: a new approach to the irrelevant speech effect. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, 22(5), 1154–1165.
- Martin, R. C., & Breedin, S. D. (1992). Dissociations between speech perception and phonological short-term memory deficits. *Cognitive Neuropsychology*, 9(6), 509–534.
- Neath, I. (2000). Modeling the effects of irrelevant speech on memory. *Psychonomic Bulletin & Review*, 7(3), 403–423.
- Neath, I., Farley, L. A., & Surprenant, A. M. (2003). Directly assessing the relationship between irrelevant speech and articulatory suppression. [comment]. *Quarterly Journal of Experimental Psychology A*, 56(8), 1269–1278, discussion 1301–6.
- Richardson, J. T. (1984). Developing the theory of working memory. *Memory & Cognition*, 12(1), 71–83.
- Salame, P., & Baddeley, A. D. (1982). Disruption of short-term memory by unattended speech: Implications for the structure of working memory. *Journal of Verbal Learning and Verbal Behavior*, 21(2), 150–164.
- Salame, P., & Baddeley, A. D. (1986). Phonological factors in STM: Similarity and the unattended speech effect. *Bulletin of the Psychonomic Society*, 24(4), 263–265.
- Salame, P., & Baddeley, A. D. (1990). The effects of irrelevant speech on immediate free recall. *Bulletin of the Psychonomic Society*, 28(6), 540–542.
- Segalowitz, S. J., & Graves, R. E. (1990). Suitability of the IBM XT, AT, and PS/2 keyboard, mouse, and game port as response devices in reaction time paradigms. *Behavior Research Methods, Instrumental and Computers*, 22, 283–289.
- Shallice, T., & Warrington, E. K. (1977). Auditory-verbal short-term memory impairment and conduction aphasia. *Brain & Language*, 4(4), 479–491.
- Sternberg, S. (1969). Memory scanning: Mental processes revealed by reaction time experiments. *American Scientist*, (57), 421–457.
- Thorn, A. S., Gathercole, S. E., & Frankish, C. R. (2002). Language familiarity effects in short-term memory: the role of output delay and long-term knowledge. *Quarterly Journal of Experimental Psychology A*, 55(4), 1363–1383.
- Welford, A. T. (1980). Choice reaction time: Basic concepts. In A. T. Welford (Ed.), *Reaction Times* (pp. 73–128). New York: Academic Press.